



OXFORD

The Transcendental Turn

edited by

SEBASTIAN GARDNER
AND MATTHEW GRIST

The Transcendental Turn

The Transcendental Turn

EDITED BY

Sebastian Gardner
and Matthew Grist

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© The several contributors 2015

The moral rights of the authors have been asserted

First Edition published in 2015

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law; by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2014941574

ISBN 978-0-19-872487-2

Printed and bound by
CPI Group (UK) Ltd, Croydon, CRO 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Acknowledgements

The papers which comprise this collection derive originally from a series of conferences organized by Mark Sacks, as part of a three-year AHRC-funded project, *Transcendental Philosophy and Naturalism*, which ran from 2005 to 2008, at the University of Essex. The companion volume deriving from the project, *Transcendental Philosophy and Naturalism*, edited by Peter Sullivan and Joel Smith (Oxford: Oxford University Press, 2011), examines the relation of transcendental philosophy to philosophical naturalism in systematic terms. The present collection is concerned with transcendental philosophy in its historical dimension.

Karl Ameriks' contribution, 'On Reconciling the Transcendental Turn with Kant's Idealism', has been published as Chapter 4 of *Kant's Elliptical Path* (Oxford: Oxford University Press, 2012). Stephen Mulhall's contribution, "'Hopelessly Strange': Bernard Williams' Portrait of Wittgenstein as a Transcendental Idealist", appeared in the *European Journal of Philosophy* 17, 2009, 386–404.

The *Transcendental Philosophy and Naturalism* project was conceived and directed by Mark Sacks (1953–2008). Mark's untimely death has been a great loss to all who knew him and to the philosophical community. This volume is dedicated to his memory.

SG, MG

Contents

<i>Notes on Contributors</i>	ix
Introduction: The Transcendental Turn <i>Sebastian Gardner</i>	1
1. From Transcendental Realism to Transcendental Idealism: The Nature and Significance of Kant's 'Transcendental Turn' <i>Henry E. Allison</i>	20
2. On Reconciling the Transcendental Turn with Kant's Idealism <i>Karl Ameriks</i>	35
3. Kant, Naturalism, and the Reach of Practical Reason <i>Paul Abela</i>	56
4. The 'Synthetic-Genetic Method' of Transcendental Philosophy: Kantian Questions/Fichtean Answers <i>Daniel Breazeale</i>	74
5. Fichte's Anti-Skeptical Programme: On the Anti-Skeptical Strategies in Fichte's Presentations of the <i>Wissenschaftslehre</i> 1794 to 1801/2 <i>Rolf-Peter Horstmann</i>	96
6. Fichte's Transcendental Ethics <i>Paul Guyer</i>	135
7. Finite and Absolute Idealism: The Transcendental and the Metaphysical Hegel <i>Robert Pippin</i>	159
8. Is Hegel's <i>Phenomenology of Spirit</i> an Essay in Transcendental Argument? <i>Stephen Houlgate</i>	173
9. Transcendental Aspects, Ontological Commitments, and Naturalistic Elements in Nietzsche's Thought <i>Béatrice Han-Pile</i>	195
10. Husserl and the Transcendental <i>Dan Zahavi</i>	228
11. Phenomenology and Transcendental Philosophy: Making Meaning Thematic <i>Steven Crowell</i>	244

12. Heidegger on Unconcealment and Correctness <i>Taylor Carman</i>	264
13. Transcendental versus Hermeneutic Phenomenology in <i>Being and Time</i> <i>Cristina Lafont</i>	278
14. Merleau-Ponty's Transcendental Theory of Perception <i>Sebastian Gardner</i>	294
15. 'Hopelessly Strange': Bernard Williams's Portrait of Wittgenstein as a Transcendental Idealist <i>Stephen Mulhall</i>	324
16. Stoic Transcendentalism and the Doctrine of <i>Oikeiosis</i> <i>Wayne M. Martin</i>	342
<i>Index</i>	369

Notes on Contributors

PAUL ABELA, Acadia University

HENRY E. ALLISON, University of California at San Diego

KARL AMERIKS, University of Notre Dame

DANIEL BREAZEALE, University of Kentucky

TAYLOR CARMAN, Barnard College, Columbia University

STEVEN CROWELL, Rice University

SEBASTIAN GARDNER, University College London

PAUL GUYER, Brown University

BÉATRICE HAN-PILE, University of Essex

ROLF-PETER HORSTMANN, Humboldt-Universität zu Berlin

STEPHEN HOULGATE, University of Warwick

CRISTINA LAFONT, Northwestern University

WAYNE M. MARTIN, University of Essex

STEPHEN MULHALL, New College, University of Oxford

ROBERT PIPPIN, University of Chicago

DAN ZAHAVI, University of Copenhagen

Introduction

The Transcendental Turn

Sebastian Gardner

Kant's influence on the history of philosophy is vast and protean. The transcendental turn denotes one of its most important forms, centred on the notion that Kant's deepest insight should not be identified with any specific epistemological or metaphysical doctrine, but rather concerns the fundamental standpoint and terms of reference of philosophical enquiry. To take the transcendental turn is not to endorse any of Kant's specific teachings, but to accept that the Copernican revolution announced in the Preface of the *Critique of Pure Reason* sets philosophy on a new footing and constitutes the proper starting point of philosophical reflection.

The question of what precisely defines the transcendental standpoint has both historical and systematic aspects. It can be asked, on the one hand, what different construals have been put on Kant's Copernican revolution and what different historical forms the transcendental project has assumed, and on the other, how transcendental philosophy, in abstraction from its historical instances, should be defined and practised. Though separable, the two questions are best taken together. In so far as historical interest has a critical dimension, it will constantly broach systematic issues, just as any convincing account of the nature of transcendental philosophy will need to take account of the historical development. The present volume examines the history of the transcendental turn with the dual aim of elucidating the thought of individual philosophers and of bringing into clearer view the existence of a transcendental tradition, intertwined with but distinguishable from other configurations in late modern philosophy. It is intended also that, by giving systematic shape to historical material, a resource will be provided for systematic reflection on transcendental philosophy.

The difficulties involved in any attempt at a definition are considerable, but there is evidently a need at the outset for a statement of the broad contours of any philosophical position that can be regarded as transcendental. Something along the following lines

may be suggested. (i) Involved in the transcendental turn in the first instance, is an acceptance of the *necessity of metaphysics*, or at any rate of some distinctively philosophical form of reflection which goes beyond empirical enquiry, along with a *criticism of earlier metaphysics* as methodologically naive, with the implication that even if correct results have been reached earlier in the history of philosophy, these have not been adequately grounded or formulated. (ii) Associated with the view that metaphysics in some form or other is required is an acceptance of the need for *final justifications* of some sort. Characteristically associated with this commitment is a recognition of the force of, and a willingness to engage with, the challenge posed by skepticism. (iii) As regards the means by which final justifications are to be provided and the form which they are to take, the Copernican revolution dictates that the foundations of knowledge derive in some manner from, or at least are to be discovered by reflection on, powers, capacities, and cognitive requirements of the *knowing subject*, and that they take the form of what Kant refers to as *conditions of possibility*, a concept that admits of various interpretations but which on all accounts contrasts with a direct appeal to matters of psychological fact. (iv) Consequent upon their separation from psychological grounds, transcendental conditions of possibility are regarded as having a rational and normative character, and as forming a systematic unity, giving rise to the notion of *transcendental logic*, which it is the task of transcendental philosophy to spell out. (v) As regards the means by which we come to grasp transcendental conditions and the basis on which we claim knowledge of their necessity, it again follows from their subjective yet non-psychological character that the transcendental turn demands a distinctive mode of *transcendental proof* or *transcendental argumentation*, either not found in the toolkit of pre-Kantian philosophy or, if present, then not explicitly theorized as such.

Other elements arguably merit inclusion in a basic statement of the transcendental turn. In particular it may be suggested that the integration of theoretical with practical philosophy, the exhibition of parallels between cognition and action, is not an accidental feature of the transcendental project.¹ While this contention is plausible and, if correct, of high importance, we should content ourselves at this point with a rough and ready, non-exhaustive sketch of the thin 'core' of the transcendental turn, since what is most wanted, for the historical purposes of this volume, is not a set of necessary and sufficient conditions but a conception of a distinctive *strategy* of maximal philosophical generality developed by Kant yet allowing itself to be executed in different ways and susceptible to revision.² According to (i)–(v), a post-Kantian development

¹ The obvious way of making the case for a principled connection of transcendental philosophy with the practical is in terms of ends: if cognition can only be understood in terms of its ends—if transcendental grounding and justification has a teleological aspect—then the conditions are in place for a single system of theoretical and practical reason. The theory of self-consciousness or subjectivity is perhaps a further candidate for inclusion in the basic transcendental agenda.

² It is relevant that Kant's own use of 'transcendental' appears to vary: compare the stricter definitions in the first *Critique* at A56/B80 and A468/B676 with the more permissive definition in the Introduction to the third *Critique* at 5: 181 and 20: 209; on this issue, see Förster 1989, and Horstmann 1989, 168ff.

incorporates the transcendental turn if it seeks to justify claims to knowledge of the necessities that govern things by reflection on and reference to the subject conceived in a non-empirical respect. To say this is not to endorse Kant's own epistemology and metaphysics, and it leaves ample room for disagreement concerning many issues: what should go under the headings of 'transcendental logic' and 'transcendental proof', what the 'systematic unity' of transcendental conditions may amount to, what counts as 'deriving from the subject', what weight should be attached to skeptical worries and the whole *questio juris*, what 'non-empirical subjectivity' might consist in, and so forth. There is a further sense in which a philosophical development, even when it is not only short of recognizably Kantian first-order doctrinal commitments but also rejects the full package of (i)–(v), may be regarded as belonging to the transcendental tradition, namely when it has come into existence by way of a critique of the Kantian transcendental strategy. Certain developments in late modern philosophy may be regarded as having worked through the transcendental option and as owing their intelligibility and suasiveness, in part at least, to their endeavour to resolve aporiae in, and 'overcome', Kant's transcendentalism. Whether we describe these as post-transcendental, or as revisionary forms of transcendentalism, will depend on the individual case.

That the transcendental turn tracks a theme in the history of philosophy will seem obvious to those accustomed to viewing the late modern development through Kantian eyes, but less so to those whose view of the philosophical geography is determined by some other trigonometric point. It is therefore worth pointing out that the notion of the transcendental turn can be employed for the purpose of historical understanding without Kantian commitment. To grant its integrity as a historical category is not to affirm that the transcendental project is in the final systematic analysis coherent. What the historical narrative may reveal, indeed, is the impossibility of designing a form of transcendental philosophy that is free from contradictions. This possibility, however disappointing it would prove, cannot be ruled out *ab initio*. Equally it should be regarded as strictly an open question how much of nineteenth- and twentieth-century philosophy is accurately described or well explained as following a transcendental agenda.³ Again it may transpire that, contrary to the expectations I have voiced, the concept of the transcendental turn fails to get significant purchase on historical figures who depart from Kantian orthodoxy, making the transcendental tradition effectively indistinguishable from the neo-Kantian tradition. This too cannot be dismissed in advance. In short, it remains to be seen what results from pursuing the hypothesis that the transcendental turn defines an autonomous vector in nineteenth- and twentieth-century philosophy. Even if the conclusions we come to are negative, they will nonetheless be instructive. The important point, and the sole

³ This collection confessedly reflects an anglophone philosophical orientation. The concept of the transcendental has figured little in mainstream analytic philosophy, and has not been especially prominent as a *historical* category in anglophone work on the post-Kantian tradition; from within other philosophical environments, the existence of a transcendental tradition may not be regarded as in need of demonstration.

premise of this collection, is that a number of the most important and creative philosophers of the last two centuries either present their thought, or allow themselves to be fruitfully interpreted, as not simply of Kantian descent but as incorporating a *metaphilosophical* insight owed to Kant and approximately captured in (i)–(v). The philosophers treated in this collection are candidates for being viewed in such terms, as the secondary literature on each attests, though their inclusion here indicates only that there is an issue to be decided concerning their relation to the transcendental turn, not which way the decision should fall. As will be seen, the contributors to this volume take different views of the transcendental credentials of the figures they discuss, as they do also, I should stress, of the general issues surrounding the concept of transcendental philosophy discussed in this Introduction.

If the transcendental turn constitutes an abiding theme in post-Kantian philosophy, then it ought to be possible to say something at a general level about why the transcendental option has seemed compelling to so many of Kant's successors and over such a long period. Many factors, of course, contribute to Kant's extraordinary reception history, but if we restrict attention to the metaphilosophical component of Kant's theoretical philosophy, then the plausible answer is that its power lies in Kant's demonstration that empiricism and rationalism in their classical forms are untenable, for reasons that go deep and pertain to their shared ultimate presuppositions, and that progress in metaphysics and vindication of the objectivity assumed in ordinary thought demand a heightened philosophical self-consciousness, which must take subjectivity, conceived however in terms that avoid the errors of Descartes and Hume, as its starting point. This positive estimate is compatible with a perception of Kant as having failed to develop fully or articulate properly the transcendental standpoint—transcendental philosophy may be considered to hold a potential which Kant has left unexploited, or to require the elimination of tensions in order for existing gains to be consolidated, or both. What follows is intended to give an idea of the sorts of issues to which the transcendental project by its nature gives rise, its constant preoccupations and characteristic trajectories. They reflect, as may be expected, perennial issues in Kant's interpretation. In order not to get drawn into the substantial interpretative issues discussed in the chapters that make up this volume, I will sketch the challenges facing and the possibilities open to transcendentalism in abstraction from the question of which historical figures they map onto.

1. Transcendental Idealism

Kant presents the transcendental turn as a turn *to* transcendental idealism. Though Kant does not dwell on the distinction, it is clear that two 'moments' are involved. Transcendental idealism comprises a first-order thesis concerning the objects, scope, and limits of human cognition. The Copernican formula, the injunction to proceed 'with the problems of metaphysics by assuming that the objects must conform to our

cognition (*Erkenntnisart*)' (Bxvi), provides its metaphilosophical and methodological complement. That they are conceptually separate, and their relation complex, becomes clear when we see how the case for transcendental idealism is made out in the first *Critique*.

If Kant's argument consisted simply and solely in a demonstration of the subjectivity of space and time, based on premises drawn from ordinary thought, then transcendental idealism might be viewed as a first-order metaphysical thesis of much broader scope than, but of the same order as, the thesis that secondary qualities are subjective. But there are compelling reasons for thinking that even Kant's arguments concerning space and time presuppose a framework which needs to be made explicit, and that in any case the Transcendental Aesthetic is merely the beginning of Kant's explication of transcendental idealism, so that it is only when the *Critique* has run its course that we come to grasp its full meaning. The following will help to bring out the point.

One familiar Kantian line of thought, associated with the Fourth Paralogism, takes its cue from skepticism. Let it be granted that the skeptic succeeds in showing the global failure of our claims to knowledge of objects, in the sense of things that exist independently of our cognizing them and which, as so existing, possess the intrinsic constitution which our cognitions represent them as having. It remains true nonetheless that we continue to seem to be presented with, and can continue to refer to, the putatively cognized and seemingly real objects which, prior to the skeptical onslaught, we took to safely meet the conditions of justified true belief. The skeptic may have razed our claims to knowledge to the ground, but as Hume stressed, skeptical arguments do not abolish the appearance of a real world and the epistemic seemings which accompany it—presented with what looks to be a blue ball, 'here is a blue ball' will continue to seem to have greater justification than its negation. And what this shows, Kant may be construed as arguing, is that we have a choice as to how we picture our epistemic situation. For what we might think is that the conception of a real object to which we were previously committed, the conception to which we subscribed in our dialogue with the skeptic, is a misconception; and that the correct conception of a real object is that of a thing which meets the full conditions of epistemic seeming, that is, a thing in so far as it agrees with our cognition.

If we take this revisionary, Copernican option, then we will be left, on the face of it, with uncertainty on two fronts. For we will not be able to claim either (a) that objects as we cognize them do not, in fact, exist just as we cognize them independently of our cognition, nor (b) that, if objects as we cognize them exist only within our cognition, there do not also exist somewhere in reality other objects which are independent of our cognition. But agnosticism regarding how things are independently of our cognition is arguably a small price to pay for security from skeptical onslaught.

The Copernican option will nonetheless have a somewhat hollow quality—as if amounting to a disguised capitulation to the skeptic—if nothing more is said about what it is for an object to 'agree with our cognition'. But here the second familiar Kantian line of thought, associated with the Transcendental Analytic, comes in. Let it

be granted that, as Kant maintains, only certain types of objects are possible objects of cognition for us.⁴ This allows itself to be developed into the stronger thesis that, whatever the objects which come our way may be like in themselves, we are bound to represent them in the terms that our power of cognition requires and makes available: such that, even if the objects of our cognition are in fact, in themselves, exactly as we represent them, it is not *because and in virtue* of their having this independent constitution that we represent them in the way that we do. If that is so, then the balance immediately tips in favour of the Copernican option: for if we are necessitated, irrespective of how things are in themselves, to represent objects as being of a certain kind, then it makes no sense to think of objects in the old way; to do so is to make the relation of knowledge contingent, an unaccountable fortuitous coincidence of how things are in themselves with how things need to be in order for us to represent them as objects—in short, just the type of ungrounded contingent relation that the skeptic made merry with. The choice we faced earlier has now become straightforward: rather than conceive objects in a way that leaves us open to skeptical doubt and allows nothing to qualify as strictly knowable, we should conceive them in a way that secures their knowability.

If this is a fair representation of the sorts of considerations involved in transcendental idealism,⁵ then Kant's thesis that the objects of our cognition are appearances is not of the same order as, say, Leibniz's thesis that bodies *qua* given in perceptual experience are merely well-founded phenomena, or Spinoza's thesis that finite particulars are modes. Something akin to an element of decision is involved in transcendental idealism—a reflective judgement concerning how it is best for us to think, how we should conceive our knowledge and its objects in light of our epistemic ambitions.⁶ This explains why we talk of the transcendental *turn*, with the implication of a shift of paradigm, in a way that we do not of a Leibnizian or Spinozistic turn. It agrees also with Kant's preference for juridical metaphors: just as a court of law decides in a case of disputed property not what is empirically the case but a matter of rightfulness, so philosophical reflection is concerned in the first place with conceptual probity, fulfilment of the subjective needs of our power of cognition, not the obtaining of metaphysical states of affairs (the latter is to be determined by way of the former).

Transcendental idealism has therefore, to state the obvious, a complex structure. Kant proceeds at distinguishable levels, on the one hand focusing on and reinterpreting the very concept of objecthood,⁷ and at another level advancing theses correlating specific formal features of objects with specific features of our mode of cognition.

⁴ 'Types' refers here to purely formal object-specifying features, prescinding from non-philosophical distinctions of ontological kinds; 'us' refers to knowers specified again in a purely formal way.

⁵ The argument just rehearsed is intended, note, only to exemplify the manner in which transcendental idealism allows itself to be argued for.

⁶ The Antinomy of Pure Reason may also be regarded as arguing for transcendental idealism in this manner.

⁷ This is the basis of Kant's retrieval of the term 'transcendental' from medieval philosophy: see Aertsen 1996, 17–24.

We may refer, I have suggested, to the former as Kant's transcendental turn, and to the latter as Kant's transcendental idealism.⁸ Importantly, the sense in which Kant's position is 'idealistic' is clearly original, for it appears to involve no claim about the ideational or mental *nature* of objects, nor is the object's relation of 'conformity to our cognition' a relation of mental dependence of any Berkeleyan or other familiar kind. To that extent, Kant appears not to 'subjectivize' reality, and the sense, whatever it may be exactly, in which his idealism contrasts with and precludes 'realism' does not seem worrisome.

Yet, as the history of Kant reception attests, this is not the end of the matter. It is true that transcendental idealism as just explicated does not comprise a judgement that knowers exercise a causal power to bestow form on objects, in the way that the doctrine of the subjectivity of colour contains the thesis that our visual apparatus projects sensational qualities onto the surfaces of physical objects. But it is not clear that *some such* first-order metaphysical, mentalistic-idealist story, whereby the mind performs an operation of form-imposition on sensational matter, will not need to be told at some point—namely, if we seek to explain *how* the conformity of objects to our mode of cognition is secured. Kant's own theory of synthesis appears to take up just this task, which arguably needs to be completed if contingency is to be eliminated from the knowledge relation.⁹ The worry that Kant's transcendental turn leads into a phenomenalist reduction casts a long shadow over the history of transcendental philosophy.

To the extent that transcendental idealism is perceived as entailing a global subjectivization of reality, there is a powerful motive for undertaking to adjust the implications of the transcendental turn as regards the opposition of realism and idealism. The correction may take various forms: it may be argued that, properly understood, the transcendental turn leads to a robust realism; or that it is neutral, without implications one way or the other; or that the idealism to which it leads is genuinely free of subjectivizing first-order metaphysical implications and can lay claim to the title of a synthetic 'realism-idealism'.

2. Conditions of Possibility

Turning to the concept of a transcendental condition, we find the same ambiguity regarding subjectivity reappearing, but also a new set of complexities.

It is not in doubt that Kant is concerned with the conditions under which we can have knowledge of objects. However, a transcendental condition cannot simply be a

⁸ From another angle they may of course be regarded as different components of transcendental idealism.

⁹ See in particular Kant's note in the Preface of *Metaphysical Foundations of Natural Science*, 4: 474–6 (Kant 2004, 10–12). Completion of the Transcendental Deduction is required for 'the question *how* the categories make such experience possible', 'the problem *how* experience is now possible by means of the categories', 'the explanation of *how* experience thereby becomes possible in the first place'. Without it, we could not avoid 'taking refuge in pre-established harmony'.

'condition of cognition' without further qualification, since formal logical requirements and empirical states of affairs may also be described as such. What makes a condition of cognition transcendental is the specific nature of the necessity that it asserts, and the basis on which it is held to be known. Now the question is this. Once it has been determined, on appropriately transcendental grounds, that a certain sensible or conceptual form—spatiality or causal order, for example—is a condition of objective cognition, has everything been said that needs to be said, or is something still owing?

One extremely broad and far-reaching issue concerns the *depth of explanation* required for transcendental purposes, the question of how far back in the order of grounds transcendental enquiry needs to go in order for its picture of our epistemic situation to be secure and transparent. The conditions which figure in the conclusions of transcendental proofs, and which specify the form of objects qua their transcendental conditioning, comprise 'front-line' conditions of possibility of objects. Once these have been determined, it can then be asked—or if it cannot, then it needs to be explained why it cannot—what underlies them. In which case transcendental philosophy will be launched into a second wave of enquiry, into 'background' transcendental conditions, conditions of the conditions which directly make objects possible. These second-level conditions, it may be argued, need to be grasped if we are not just to know *which* necessities govern our cognition, but also to have *insight* into them.

The question of whether it is necessary to extend transcendental enquiry into something closer to metaphysical explanation sharply divides Kant's successors, and it is not hard to see how the case can be made either way. On the one hand, it would seem part of the original sense of the transcendental turn, a concomitant of its critique of pre-Critical metaphysics, that knowledge of front-line transcendental conditions be deemed sufficient. Thus it may be argued that, once we have grasped the functional congruence of certain objectual forms with our mode of cognition, the relevant transcendental necessity *has* been explained; to ask for explanation in any other, stronger sense is to abandon the principles of transcendentalism. On the other hand, it seems that nothing in the concept of transcendental grounding forbids our asking, in accordance with the principle of sufficient reason, *why* our cognition should be such that it necessitates certain objectual forms, and arguably the ambition of providing final justifications warrants this further enquiry.

The other issue that arises is narrower, and again marks a sharp division in construals of the transcendental turn, related to the one just described.

On one view, purely *epistemological* characterizations are *sufficient* to give the full essence of transcendental conditions. For instance: transcendental conditions may be identified with conditions under which beliefs are justified or become candidates for justification.¹⁰ This gives transcendental philosophy a ready intelligibility and

¹⁰ Paul Guyer claims that Kant's arguments yield 'an epistemological model of the confirmation of beliefs,' 'the basic framework for the justification of beliefs': Kantian conditions of possibility are 'principles which would have to be appealed to in the justification of empirical claims to knowledge,' 'conditions for verifying or confirming empirical judgements' (Guyer 1987, 258–9 and 304). In stating and defending his position,

accords with its anti-skeptical thrust, and it has been the dominant view at least since neo-Kantianism. As might have been expected in view of the neo-Kantians' ambition to steer a course between materialism's rejection of the *a priori*, which destroys objective knowledge, and German idealism's perceived regression from transcendental philosophy to dogmatic speculation, which jeopardizes the authority of philosophical reason, we find in Hermann Cohen and his successors a series of explicit attempts to put the transcendental turn in focus. The following passage from Ernst Cassirer provides a clear and precise statement of how the transcendental turn, epistemologically conceived, may be held to fulfil the aim of giving transcendental conditions enough reality of the right kind:

The essential characteristic of Kant's transcendental method consists in the fact that it operates not in the realm of empirically real things or events, but purely and exclusively in the realm of *truths* and their ideal mode of validity. Here it is asked, not what a thing is or is not, nor what properties it possesses or what effects proceed from it, but what necessary and universal judgements precede all others. For according to Kant 'transcendental' applies to those cognitions which are concerned not so much with objects as with our mode of cognition of objects, in so far as these are to be possible *a priori*. Not the structure of objects, but the structure of judgements of empirical knowledge, by virtue of which alone objects can be known and given, stands first in line;—it is not the final ground of being which is to be disclosed, but the principles and conditions of mathematical certainty on the one side and of empirical certainty on the other, which are to be determined. In this way the pure forms of intuition and fundamental concepts of the pure understanding are discovered. We comprehend the ground of their validity when we grasp that they are not borrowed from objects but are themselves the original functions of objectual connection. Thus transcendental logic as a pure all-comprehension (*Inbegriff*) of *rules* becomes the norm for all statements about *being*. For the being of which we have in general systematic knowledge can be grasped by us only in *judgement*;—the truth of derived judgements presupposes that of fundamental judgements. Once these have been reached, it is impossible for them to be grounded in anything further, and no *being*, whether psychic or physical, empirical or metaphysical, can be designated as that on which they depend or that through which they are justified. On the contrary, 'deduction' must proceed here in the opposite direction: it must be shown that all *claims* regarding any such being already imply the validity and truth of those judgements and so that any attempted derivation of them from things would be circular.

Such a *circulus vitiosus* is what necessarily results as soon as the question of Critical philosophy is taken out of its own proper domain;—as soon as we ask not about an ideal 'dependence' of truths, but about a real dependence of the subject on the object or vice versa. For whatever may be the ultimate answer to this question, the very posing of it displaces the original *logical* relation of validity, the relation of super- or subordination within the system of knowledge, into a *meta-physical* relation of 'inherence' or causality.¹¹

Guyer contrasts his reading of transcendental conditions with the view that they constitute psychological factors in the production of belief. Cf. Henry Allison's denial that Kant (in the *Refutation of Idealism*) is concerned with the 'conditions of the justification or verification' of knowledge claims (Allison 1983, 297).

¹¹ Cassirer 1923, 427–8 (my translation).

The central distinction exploited by Cassirer, we see, is that of purely normative from existentially committed claims. Transcendental conditions on his account are necessarily without ontological commitment, and this feature is more fundamental than their non-empirical character. As Cassirer argues the point, any ontological component or grounding of transcendental conditions would in the first place be redundant—it could contribute nothing to the justification of principles without having itself been validated by those principles—and secondly it would abort the task of justification by substituting non-normative description for transcendental analysis. Immunity to skeptical attack, it may also be argued, imposes the same requirement: if transcendental knowledge is premised on claims to knowledge of anything existent, then it is open to challenge concerning our presumed epistemic access to the entity in question (no less so in the case of mental ‘faculties’ than in that of external objects).

Cassirer’s account may be probed on several fronts. In particular, it may be asked how the spheres of the ‘ideal’ (normative) and the ‘real’ (factual) are connected, as it seems they must be, if the former is to determine the latter, as Cassirer appears to affirm. As he explains matters, the ‘fundamental judgements’ which constitute the presuppositions of objective knowledge refer us to the pure forms of sensibility and understanding, which constitute ‘the original functions of objectual connection.’ There is therefore, it would seem, a derivation of the ‘real’ from the ‘ideal’, or an informing of the former by the latter, and the relation which this presupposes must presumably itself be real; making it hard to understand how transcendental grounds can be of a purely epistemological nature.

The difficulty facing Cassirer points to a general reason for considering the epistemological characterization of transcendental conditions too minimal. Kant is explicit that conditions of possibility are not just conditions of the possibility of judgements or claims to knowledge, but also conditions of *things*, the *objects* of judgements. For example:

we can readily grasp the possibility of community (of substances as appearances) if we represent them in space, thus in outer intuition. For this already contains . . . conditions of the possibility of the real (in effect and countereffect, thus in community).—It can just as easily be established that the possibility of things as *magnitudes* . . . can also be exhibited only in outer intuition. (B293)

It may be replied that talk of conditions of possibility of substances and their real relations, of things and their quantitative properties, is simply shorthand for talk of those items qua judged and known, that is, for conditions of knowledge. But if that is so, then the following well-known, key statement from the beginning of the *Transcendental Analytic*, which Kant appears to regard as substantial, reduces to a tautology: ‘The conditions of the *possibility of experience* [*Erfahrung*, identified by Kant with *empirisches Erkenntniß*, empirical cognition, B218] in general are at the same time conditions of the *possibility of the objects of experience*, and on this account have objective validity’ (A158/B197). The notion of a condition of possibility, applied to objects, implies

a transitive relation of making-possible, different in kind from the normative relations involved in knowledge, and underpinning them. It may also be pointed out that a reduction to epistemology, though it may be *suggested* by the Copernican formula, is not what it strictly *asserts*: to consider objects in relation to our mode of cognition is still to consider *objects*, not just their representations, and thus, arguably, to continue to treat of things qua their being. So it may be argued that there is as much reason to identify transcendental philosophy with a form of ontological enquiry distinguished by its use of the form of cognition as an ontological principle, as with a supplanting of ontology by epistemology.¹²

Following the historical trail leads to the following suggestion for developing the ontological view. The notion of a condition of possibility carries forward a form of philosophical explanation found in Kant's predecessors, belonging to the rationalist package elaborated out of Leibniz by above all Christian Wolff, and re-employed with amendments by later eighteenth-century German philosophers.¹³ On this conception (compressing into one the views of several authors, and simplifying massively) all truths, for example those of geometry or natural science, imply the existence of entities beyond the objects of experience. These further entities—'Realitäten', in Wolff's terminology—are known through clear and distinct ideas or 'real definitions', *Realdefinitionen*, and provide the determining ground and essences of the entities given to us in experience. *Realitäten* themselves are not given in sense experience, their nature being purely intellectual, and their ontological home is in God's mind: they constitute as it were God's thoughts. They comprise the sphere of extra-logical possibility, for nothing is possible which is not composed out of them as the basic material of finite determinate being. Kant's recognition of the coherence and

¹² The 'ontologische Kantinterpretation' of German scholarship in the 1920s and 1930s, associated with Max Wundt, Heinz Heimsoeth, and Martin Heidegger, among others, set itself against the epistemological orientation of neo-Kantianism. The considerations which it chiefly cites in its support concern the practical fulfilment of reason, the positive function of the Transcendental Dialectic, and Kant's late talk of a 'practico-dogmatic' metaphysics. But the case for an ontological reading can also be made—Heidegger does so—on grounds of the Aesthetic and Analytic of the first *Critique* alone. Heidegger refers to Kant's identification of ontology with transcendental philosophy in the late *What Real Progress?* text (20: 260, 20: 263, 20: 281).

¹³ Moses Mendelssohn provides a clear example in 'On Evidence in the Metaphysical Sciences' (1764): 'Of the properties of things outside us, we never know with convincing certainty whether they are realities (*Realitäten*) or mere appearances (*Erscheinungen*) and, at bottom, depend upon negations; indeed, in the case of some of them, we have reason to believe that they are mere appearances. Thus we can ascribe none of these properties (*Eigenschaften*) to the Supreme Being and must absolutely deny him some of them. Belonging to the latter group are all *qualitates sensibiles* that we have reason to believe are not to be found outside us as they seem to us thanks to our sensuous, limited knowledge and that, therefore, are not realities (*Realitäten*). This inference can also be inverted. What does not belong to the Supreme Being cannot be a reality (*Realität*) since all possible realities (*Realitäten*) are his to the highest degree. From this it follows naturally that extension, movement and color are mere appearances and not realities (*Realitäten*). For, were they realities (*Realitäten*), then they would have to be ascribed to the Supreme Being' (Mendelssohn 1997, 290). The concept goes back to Scotus, who recognizes the need for a type of distinction midway between real distinctions and conceptual distinctions, which he calls a formal distinction: what it distinguishes are not things (*res*) but 'formalities' or 'realities' (e.g. God's will and God's intellect).

appeal of this rationalist structure is testified by his detailed and precise restatement of it in the Ideal of Pure Reason, where it supplies the basis for his transcendental account of the Idea of God.

Now Kant of course rejects the notion that the *Realitäten* relevant to human knowledge are mind-independent, metaphysically real entities, but he does not reject the associated form of philosophical explanation: he agrees that the things that we are capable of knowing cannot derive their reality from themselves, and that something must give them reality and make them possible. This something, Kant maintains, is the form or act of cognition. Kant has therefore subjectivized the rationalist apparatus, by accepting that there are indeed *Realitäten* in the sense of things which play the metaphysical role of giving the object determinate existence, but identifying them with our power and activity of cognition: human cognition occupies the same position in relation to phenomena as the Transcendental Ideal does in relation to noumena. Cognition is thus not fundamentally a matter of 'being justified in making such and such a claim', but rather 'that by virtue or on account of which there come to be objects of such and such a kind'; and it is because cognition supplies a metaphysical ground that the subject is able to make justified claims to knowledge of objects. Whereas the epistemological reading reduces conditions of possibility to non-metaphysical conditions for knowledge, the present reading treats knowledge, in the sense of the forms and activity of cognizing, as itself a metaphysical condition for the existence of the object as cognized.

It appears that, in seeking to correct the overly austere purely epistemological reading, the transcendental turn has been brought full circle back to metaphysics. Such a conclusion may strike us as forced and improbable. Certainly the notion that transcendental philosophy amounts to, as it were, 'metaphysics by epistemological means' is not lucid. But the challenge remains to explicate Kant's claim at A158/B197 concerning the parallelism or identity of the conditions of possibility of empirical knowledge and those of its objects. What has in any case been brought to our attention is the intricate interplay of epistemological and metaphysical/ontological factors in transcendental explanation, and the difficulty of understanding how it relates to the distinction between the order of things and the order of knowledge: the relation of transcendental conditioning seems not to allow itself to be factorized in such terms.¹⁴ If the transcendental turn is adequately characterized as an epistemological turn, then it is because cognition turns out to be not quite what we previously thought. The puzzle arising from transcendental philosophy's apparent fusion of epistemological and ontological grounds surfaces repeatedly throughout its history.

¹⁴ The passage from B293 quoted earlier illustrates the point. Kant tells us that the possibility of things as magnitudes can be exhibited (*kann dargelegt werden*) in (and only in) outer sense. How is the exhibitability relation to be parsed: is outer sense constitutive of existence as magnitude, or is it only the unique means of epistemic access to such existents? The puzzle is reflected in several long-running issues in Kant interpretation; e.g. concerning how the distinction is to be drawn between space and the representation of space, and between appearances and representations.

3. The Transcendental Standpoint

The concept of ‘standpoint’ or ‘point of view’ recurs in transcendental discourse. It is employed at several points in formulating the doctrine of transcendental idealism: the ‘human standpoint’ to which all knowable objects are indexed is held to be defined by the spatio-temporality of our sensibility; and the standpoint from which nature is judged to be ‘ideal’ is that of transcendental reflection, while it is judged ‘real’ from the non-transcendental standpoint of natural consciousness.¹⁵ And it appears again in the self-description of transcendental idealism, which presents itself as discontinuous with all earlier history of philosophy, the standpoint of which has been (tacitly, unconsciously) ‘transcendental realist’.

It is not surprising, given the weight that it bears, that a number of perennial issues in transcendental philosophy should be bound up with the concept of a transcendental standpoint, and that it should have been central to one notable early attempt to state with greater exactness what distinguishes Kant’s transcendentalism. Jakob Sigismund Beck, a contemporary of Kant’s, edited a multi-volume set of extracts from his Critical writings, the third and concluding volume of which, published in 1796, he describes as ‘presenting the only possible standpoint from which the Critical philosophy is to be judged’. The thought that drives Beck, and many later engagements with the transcendental turn, is that the various puzzles and opacities of Kantian philosophy, whereby it opens itself to criticism, are due to an inadequate appreciation of the ‘interiority’ of transcendental philosophy—a failure to hold fast to the defining self-conscious commitment of the transcendental philosopher to exclude all realist-cum-dogmatic prejudice, and to consider matters in exclusively transcendental terms.

Without entering far into the details of his account, it is instructive, as an early anticipation of future transcendental developments and an index of the questions that it generates, to see how Beck develops the standpoint doctrine.¹⁶ In the first *Critique*, Beck suggests, Kant is hampered in his exposition of his ideas by the need to lead the reader gradually up to the relevant standpoint, which involves the acquisition of new concepts, which is dependent in turn on an extra-discursive factor. This factor Beck identifies with the *activity* of the understanding ‘through which it *originally* creates for itself the concept of an object and produces the “*I think an object*”’. This so-called ‘*original attribution*’, the setting-up or instituting of the situation in which an object can be presented with a subject, Beck compares to the geometer’s postulate ‘Conceive of space!’ Everything depends, Beck emphasizes, on our getting ourselves *into* this activity, on our identification with a vantage point which, once occupied, generates its own framework, i.e. the panoply of Kantian concepts, just as the geometer who has conceived of space can then construct determinate figures:

I say that one can only have a full understanding of all these things [viz. central claims of the *Critique of Pure Reason*], and even of the discursive concept ‘possibility of experience’ itself,

¹⁵ This notion receives a detailed exposition (and criticism) in Moore 1997.

¹⁶ See Beck’s letter to Kant, 17 June 1794, in Kant 1999, 479–81 (no. 630, 11: 508–11).

when one has fully mastered this standpoint. So long as one still thinks of this 'possibility of experience' purely discursively and does not follow the original attributive activity in just such an attribution as this, one has insight into virtually nothing, having merely substituted one incomprehensible thing for another.¹⁷

Whether or not Beck is right to depict understanding Kantian thought as a strictly all or nothing matter, and whether or not its aporiae truly disappear as soon as we occupy his standpoint, Beck is surely on the right path: some self-conscious rupture with ordinary thought is required, if the transcendental turn is to reveal itself as a genuine meta-philosophical innovation, not just a constructive compromise between empiricism and rationalism which allows each to assume more sophisticated forms. Nothing less is implied by the idea that to take the transcendental turn is to embark upon a change of paradigm. A *prima facie* implication of this view worth noting is that 'transcendental arguments' cannot be singled out purely on the basis of their form and the characteristic topic of their premises; in other words, in order to *follow* such arguments or proofs, attention to more than the deductive structure of their discursive presentation is required.¹⁸

This alerts us to yet another set of issues, revolving around the question of whether the transcendental turn can vindicate itself conclusively. If a new set of rules is to govern philosophical enquiry, this raises in acute form the question of proof and justification. If the turn is not the result of a linear argument, to what extent can its claim to rational necessity avoid circularity? And if it involves the creation of a new form of intelligibility, as Beck for one seems to suggest, does it not then involve ultimately something akin to a Kuhnian gestalt switch?

A revolution is a historical change in which a new regime claims authority on a basis unacceptable to the old. Revolutions in philosophy are thus exposed to the charge of having simply deserted the old questions in place of answering them, of having changed the topic and merely *stipulated* what is and is not allowed to count as a cogent philosophical issue. Hence the charge levelled by Kant's contemporary opponents: the Copernican revolution demands its own dogmatic foundation, so Kant's own philosophy fails to be Critical. And arguably this combination of stipulation and circularity is brought out by Beck's attempt to clarify Kant's transcendentalism: unless we *already* know the truth of transcendental idealism (and hence have already understood it, *pace* Beck), how can we *know* that the standpoint of the 'original attribution' is truth-conducive and not illusory, a standpoint of mere fictional construction?

Transcendental thought represents itself as located on one side of a line, on the other side of which are ranged the various manifestations of transcendental realism. These include the conflicting dogmatic positions claiming knowledge of things in themselves, skeptical responses to these, and the naive pre-philosophical standpoint which

¹⁷ In Kant 1999, 480 (11: 510).

¹⁸ See Sacks 2005.

sponsors dogmatic metaphysics. The transcendental philosopher claims to be able to make sense of transcendental realism in a way that it cannot make sense of itself, releasing it from its conundrums, which is why, of the two standpoints, it is that of transcendental idealism which is to be preferred.

But even if it is true that transcendental idealism enjoys a consistency lacking in transcendental realism—that is, even if there were (contrary to historical fact) universal agreement among transcendental idealists concerning the implications of the doctrine, in contrast to the disagreements raging in the transcendental realist camp—it does not follow directly that transcendental idealism is the *right* standpoint. An impartial observer, undecided between transcendental idealism and transcendental realism, may reasonably hesitate, for it may be that transcendental idealism brings hostilities to end on false terms, cutting the Gordian knot instead of untying it in the way that reason requires.¹⁹ Transcendental idealists may be justified in their *own* terms in believing that they are able to *see through* the confusions of transcendental realism, but the presuppositions of this diagnosis are ones that the transcendental realist is bound to reject; and since the realist ‘analysand’ cannot accept the interpretations of the idealist ‘analyst’, the therapy cannot work.

Suppose that, for the reasons just given, a clean supersession of transcendental realism is not forthcoming. What then suggests itself as a task for transcendental

¹⁹ Here is one way the criticism might be developed, which goes back to the earlier point concerning transcendental philosophy’s ‘fusion’ of epistemology and metaphysics. The transcendental philosopher declares that matters of ontology are to be approached by way of reflection on our representations and cognitive powers. Yet transcendental enquiry is not interested in representations to the *exclusion* of objects. Rather the intention is to redetermine objects *by way of* representations. This then raises the question: At what point, if any, are we entitled by transcendental lights to infer ‘*p*’ from ‘It is necessary that we represent *p*’? The transcendentalist answer, it would seem, can only be that we are *never* entitled to *absolutely* discharge the subjective-representational prefix, to disembody *p* and transfer necessity altogether to *what is thought* as distinct from the *thinking* of it. The transcendental realist accordingly rejoins: The task of philosophy is *defined* by the challenge to show that there *are* conditions under which inferences to fully extra-representational necessities are legitimate; if this is not broached, then the task has been abandoned. Moreover, the idealist’s denial of the possibility of such an inference evinces dogmatism. The transcendental idealist appears captivated by a picture of ‘being confined within representations’, in the same way that Berkeley appears captivated by the thought that whatever he imagines occurs within an act of imagining. And this is confused: just as, in the Berkeley case, the act of imagining is not included within the *content* of what is imagined, so when I arrive through transcendental proof at the judgement ‘There must be substance’, I am not making a judgement about my own acts and practices of conceptualization and inference; I am not, in that judgement, *resolving* to use a concept or *targeting* my own cognitive states. Cassirer’s statement in the quotation, ‘the being of which we have in general systematic knowledge can be grasped by us only in *judgement*’, also exemplifies the confusion: to be sure, judgement (and not being) is what figures in contexts of justification (we justify claims, not things), but the judgements *for* which justification is sought are precisely claims concerning what is taken to have *being* independent of the making of the judgement. Hence the paradoxical nature of Kant’s empirical realism: ordinary claims about empirically real objects do not make reference to their own claim-making.

This sort of realist counter-blast is characteristic of Kant’s contemporary critic F. H. Jacobi. For a sophisticated recent realist critique of transcendentalism, roughly along these lines, see Moore 1997, esp. ch. 6, and Moore 2012, ch. 5.

philosophy is a conciliatory project, a demonstration that the transcendental standpoint is able to recuperate some of the territory that it left behind in its initial moment of self-formation. *Rapprochement* may be sought either with the metaphysical tradition or with naturalism. The former belongs to the nineteenth century, while the latter is of course the option of our day. The leading idea for the former is that transcendental philosophy can re-engage with the old questions and the sphere of speculative metaphysics be recouped on transcendental grounds: 'background' transcendental conditions, it may be argued, have the same reach as the intuitions of reason to which early modern rationalism laid claim. In the case of proposed unifications with naturalism, the idea will be that transcendental conditions, as well as being discoverable a priori, can also be accessed in a posteriori form, allowing philosophical reflection to join itself with the natural and human sciences. The most pressing question in this context is whether allowing transcendental conditions a double identity undermines the necessity which is an integral part of what gives them their philosophical interest in the first place.²⁰

* * *

The organization of the collection is straightforward: with one exception, the chapters follow the historical time-line.

The first three, by Henry Allison, Karl Ameriks, and Paul Abela, deal with Kant. Allison focuses on Kant's concept of the transcendental and articulates a deflationary, non-metaphysical interpretation thereof, which provides the basis for an elucidation of the opposition of transcendental idealism and transcendental realism. Ameriks by contrast develops an ontological interpretation of Kant's idealism, which he characterizes as non-subjectivist and 'moderate', and which, Ameriks argues, addresses the concerns that lie behind the non-metaphysical reading. Abela, turning to practical reason, argues that if Kant's moral theory is interpreted, as it ought to be, as having its centre in the concept of the highest good, then the distance separating Kantian ethics from a naturalistic view of the moral life is narrowed considerably; framing the good life in transcendentalist terms does not, Abela argues, displace naturalistic categories.

Fichte is discussed by Daniel Breazeale, Rolf-Peter Horstmann, and Paul Guyer. Again the first two chapters focus on the theoretical and the third on the practical. Breazeale opens with a consideration of how Kant leaves matters regarding the status and justification of transcendental claims, and proceeds to a study of the transcendental method that Fichte developed in his early *Wissenschaftslehre*, which is defined by its 'genetic', 'dialectical', and 'synthetic' dimensions, and involves the transcendental philosopher in acts of mental construction. Horstmann considers how Fichte attempts to defend on transcendental grounds the possibility of knowledge against

²⁰ See esp. Sacks 2003, ch. 6. Other writings relevant to the question of what constitutes and defines the transcendental turn include Hartmann 1966; Krings 1984; Carr 1999; and Bell 2001.

skeptical objections: sharing the dissatisfaction of many of his contemporaries with Kant's achievement in this regard, Fichte shifts, Horstmann argues, from a 'justification' strategy to a 'grounding' strategy, in which a special role is occupied by the theory of self-consciousness, and which Horstmann traces through to Fichte's 1801/2 *Wissenschaftslehre*. Guyer focuses on Fichte's attempt, in his 1798 *System of Ethics*, to improve on Kant's own attempt to ground moral philosophy within the transcendental framework. Three problems in particular are to the fore: the transcendental deduction of the fundamental principle of morality, the element of feeling in moral awareness, and the problem of freedom posed by the fact that human beings do not always act as morality demands.

Robert Pippin and Stephen Houlgate offer contrasting considerations of Hegel's relation to transcendentalism and thus of Hegel's project as a whole. Pippin, here attending to the notions of metaphysics and logic in Hegel, and to Hegel's view of the relation between the conceptual and the sensible, defends an interpretation of Hegel's idealism as a development of Kant's, but one which is neither 'merely' transcendental nor such as to entail a return to 'substantialist' pre-Critical metaphysics. Houlgate contests Charles Taylor's account of Hegel's *Phenomenology of Spirit* as 'an essay in transcendental argument': in fact, according to Houlgate, Hegel criticizes the transcendental approach to natural consciousness; and what is offered in the *Phenomenology* is not a philosophical theory of consciousness and its conditions, but a rigorously phenomenological study of what emerges *in* and *for* consciousness.

Nietzsche is discussed by Béatrice Han-Pile, who argues that the interpretations proposed in the literature of Nietzsche as either a transcendental philosopher or a naturalist are too simple, while the common proto-deconstructionist reading misconstrues Nietzsche's error theory. Rather, Han-Pile argues, Nietzsche efforts are directed at an overcoming of the opposition between transcendental philosophy and naturalism; though denying the possibility of genuine a priori knowledge, Nietzsche remains close to transcendental idealism.

The phenomenologists are discussed by Dan Zahavi, Steven Crowell, Taylor Carman, Cristina Lafont, and Sebastian Gardner. Zahavi focuses on Husserl's use of the term 'transcendental', methodology, and relation to Kant, arguing that Husserlian phenomenology certainly deserves to be classified as a form of transcendental philosophy, though it operates with a conception of the transcendental that, owing to Husserl's intersubjective transformation of subjectivity, differs in important respects from the traditional one. Crowell, focusing on Husserl and Heidegger, argues all phenomenology to be transcendental, insofar as it makes *meaning* the primary topic of philosophy: rather than restricting transcendental inquiry to knowledge claims, phenomenology expands the scope of transcendental philosophy to all experience, investigating its normative structure; the subject's responsiveness to norms, Crowell argues, is revealed by phenomenology as the key to a transcendental account of meaning or intentionality. Carman examines the theme of truth in *Being and Time* and later writings of Heidegger's. Heidegger's account of truth as uncovering is intended, Carman

argues, as a phenomenological characterization of the experiences and practices that warrant our deeming something true; and this conception of truth as ontologically foundational, a condition for the intelligibility of practical and theoretical attitudes, shows Heidegger to remain rooted in the tradition of transcendental philosophy. Lafont analyses conflicting tendencies in Heidegger's *Being and Time*: Heidegger aims to single out the essential structures of human existence in the ahistorical mode of transcendental philosophy, yet his hermeneutical phenomenology reveals our radical facticity, historicity and situatedness. The solution, Lafont argues, is to view hermeneutic phenomenology as a variety of transcendental philosophy: hermeneutic conditions of understanding are the ultimate transcendental conditions, which consequently forego their claim to universal validity. Gardner argues that Merleau-Ponty's theory of perception should be understood as belonging to transcendental philosophy, and as committed to transcendental idealism, for which reason it is not comparable to apparently similar theories in analytic philosophy of mind: Merleau-Ponty's interest in scientific psychology is subordinated, Gardner argues, to his strategy of revealing 'pre-objectivity' as a final transcendental condition.

Stephen Mulhall provides a critical examination of Bernard Williams's highly influential interpretation of Wittgenstein as a transcendental idealist, focusing on Wittgenstein's attitude to empirical science, conception of (grammatical) limits, and use of the first person plural; misunderstanding of Wittgenstein's position and attitude in these contexts, Mulhall argues, underlies Williams's misconstrual. The concluding chapter by Wayne Martin challenges the historical identification of transcendentalism with the Kantian legacy: traces of the transcendental strategy, Martin argues, can be found already among the ancients, in the Stoic doctrine of *oikeiosis*, a distinctive form of self-consciousness which, Martin shows, serves as a condition of the possibility of intentionally determinate perceptions and desires.

It is obvious, but all the same should be acknowledged, that the selection of historical figures who receive discussion in this volume is not beyond challenge and that the list of those who have been omitted (for simple reasons of space) but whose relation to the transcendental tradition is either clear or in need of elucidation—Schelling, Schopenhauer, Peirce, Bergson, Sartre, Levinas, Adorno, Foucault, among others—is long and distinguished.

References

- Aertsen, Jan A. (1996), *Medieval Philosophy and the Transcendentals: The Case of Thomas Aquinas* (Leiden: E. J. Brill).
- Allison, Henry (1983), *Kant's Transcendental Idealism: An Interpretation and Defense*, 1st edn (New Haven: Yale University Press).
- Bell, David (2001), 'Is Empirical Realism Compatible with Transcendental Idealism?', in Ralph Schumacher (ed.), *Idealismus als Theorie der Repräsentation?* (Paderborn: Mentis), 167–79.

- Carr, David (1999), *The Paradox of Subjectivity: The Self in the Transcendental Tradition* (Oxford: Oxford University Press).
- Cassirer, Ernst (1923), *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit* (Berlin: Verlag Bruno Cassirer).
- Förster, Eckart (1989), 'Kant's Notion of Philosophy', *The Monist*, 72, 285–304.
- Guyer, Paul (1987), *Kant and the Claims to Knowledge* (Cambridge: Cambridge University Press).
- Hartmann, Klaus (1966), 'On Taking the Transcendental Turn', *Review of Metaphysics*, 20, 223–49; reprinted in *Studies in Foundational Philosophy* (Amsterdam: Editions Rodopi, 1988).
- Horstmann, Rolf-Peter (1989), 'Why Must There Be a Transcendental Deduction in Kant's Critique of Judgement?', in Eckart Förster (ed.), *Kant's Transcendental Deductions: The Three 'Critiques' and the 'Opus Postumum'* (Stanford, CA: Stanford University Press), 157–76.
- Kant, Immanuel (1998), *Critique of Pure Reason*, tr. and ed. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press).
- Kant, Immanuel (1999), *Correspondence*, tr. and ed. Arnulf Zweig (Cambridge: Cambridge University Press).
- Kant, Immanuel (2004), *Metaphysical Foundations of Natural Science*, tr. and ed. Michael Friedman (Cambridge: Cambridge University Press).
- Krings, Hermann (1984), 'Knowing and Thinking: On the Structure and History of the Transcendental Method in Philosophy', tr. Helga Schmettau and Darrell E. Christensen, in Darrell E. Christensen et al. (eds), *Contemporary German Philosophy*, iv (University Park, PA: Pennsylvania State University Press), 206–22.
- Mendelssohn, Moses (1997), 'On Evidence in the Metaphysical Sciences' (1764), in *Philosophical Writings*, ed. Daniel O. Dahlstrom (Cambridge: Cambridge University Press), 251–306.
- Moore, A. W. (1997), *Points of View* (Oxford: Oxford University Press).
- Moore, A. W. (2012), *The Evolution of Modern Metaphysics: Making Sense of Things* (Cambridge: Cambridge University Press).
- Sacks, Mark (2003), *Objectivity and Insight* (Oxford: Oxford University Press).
- Sacks, Mark (2005), 'The Nature of Transcendental Arguments', *International Journal of Philosophical Studies*, 13, 439–60.

1

From Transcendental Realism to Transcendental Idealism

The Nature and Significance of Kant's 'Transcendental Turn'

Henry E. Allison

In a conversation with Schiller in 1803, Madame de Staël is reported to have asked about the meaning of 'transcendental' and was told that whoever understands this term understands the Kantian writings.¹ Although this is not strictly true, I believe it to be the case that whoever wishes to understand Kant must understand the ways in which he uses this term. The present chapter is an attempt to contribute to such an understanding. Its thesis is that Kant's so-called 'transcendental turn' is best understood as a turn *from* one form of transcendentalism to another rather than simply as a turn *to* the transcendental. Kant calls the form of transcendentalism from which he turns 'transcendental realism' and the form to which he turns 'transcendental idealism'.

The discussion is divided into four parts. The first provides a brief account of Kant's use of the term 'transcendental' in the *Critique*. It claims that he uses it in at least two distinct senses, one of which reflects its traditional ontological use as referring to what pertains to things or objects in general, and the other his distinctive 'critical' concern with the determination of the conditions and limits of cognition. By focusing on the contrast between a transcendental and a merely empirical realism and appealing to the inseparability of the latter from transcendental idealism, I argue in the second part that the turn from transcendental realism to this idealism is to be understood as a deflationary, critical move regarding the scope of spatiotemporal predicates rather than

¹ The information comes from the entry 'Transzendental; Transzendentalphilosophie' by Norbert Hinske, *Historisches Wörterbuch der Philosophie* (Darmstadt: Wissenschaftliche Buchgesellschaft, 1998), x. 1358.

as the dogmatic adoption of a subjectivistic metaphysics. The third and fourth parts then attempt to illustrate and support this claim by showing how the contrast between the two generic forms of transcendentalism functions in the Aesthetic and Dialectic respectively.

I

Although Kant's understanding and use of the term 'transcendental' is an extremely complex and controversial matter to which I can hardly do justice here, I do hope to say enough to clarify my thesis concerning the nature of Kant's transcendental turn. I shall begin with a consideration of two texts in the *Critique* where Kant offers what amounts to definitions of 'transcendental'. As we shall see, rather than being devoted exclusively to one of the senses I have distinguished, each definition expresses something of the ambiguity and tension of the Kantian conception.

The first text is in the Introduction, where Kant provides stipulative definitions of the term. I use the plural advisedly because there is a subtle, yet significant, difference between the definitions offered in the two editions. In the first edition, Kant states: 'I call all cognition *transcendental* that is occupied not so much with objects but rather with our *a priori* concepts of objects in general' (A11–12).² In the second edition, this is changed to read: 'I call all cognition transcendental that is occupied not so much with objects but rather with our mode of cognition of objects insofar as this is to be possible *a priori*' (B25).³

Setting aside the obvious ambiguity in the expression 'not so much with objects' (*nicht sowohl mit Gegenständen, sondern mit*), which suggests that transcendental

² All references to the *Critique of Pure Reason* are to the standard A/B pagination of the 1st and 2nd edns and cite the translation of Paul Guyer and Allen Wood, *The Cambridge Edition of the Works of Immanuel Kant* (Cambridge: Cambridge University Press, 1998). References to other works of Kant are to the volume and page of *Kants gesammelte Schriften*, ed. Deutschen (formerly Königlich Preussischen) Akademie der Wissenschaften, 29 vols. (Berlin: Walter de Gruyter, and predecessors, 1902–). Citations from the translation of Kant's Inaugural Dissertation (abbreviated as ID) are to the translation by David Walford, Cambridge Edition, *Theoretical Philosophy 1755–1770*, tr. and ed. David Walford in collaboration with Ralf Meerbote (1992); from the *Prolegomena to any Future Metaphysics* (abbreviated as *Prol.*) are to the Gary Hatfield translation, Cambridge Edition, *Theoretical Philosophy After 1781*, ed. Henry Allison and Peter Heath (2001).

³ Although these definitions have been frequently discussed in the German, philologically oriented literature, the most thorough treatment of the subject is by Tilmann Pinder, 'Kant's Begriff der transzendentalen Erkenntnis', *Kant-Studien*, 77 (1986), 1–40. According to Pinder, in the A version Kant is trying to indicate that the central focus of transcendental cognition and, therefore, of the *Critique* itself will be on our *a priori* concepts of objects rather than on objects (or things) themselves, which would characterize the ontological approach. Since a concern with such concepts involves also one with the objects (if any) supposedly falling under them, it will be concerned (albeit indirectly) with the latter as well. Hence, Kant's use of the 'not so much . . . but rather' (*nicht sowohl . . . sondern*) locution. By contrast, in the B version, Pinder thinks that Kant's focus has shifted to a more narrow concern with the possibility of synthetic *a priori* judgments, which reflects the central concern of the *Prolegomena*. Since the details of this shift, as important as they may be for an understanding of the development of Kant's thought, are not directly relevant to the concern of this chapter, I have attempted to provide a characterization of Kant's definitions that covers both versions.

cognition is in some way concerned with objects, albeit not primarily or directly, the main point is that the second definition, much more clearly than the first, indicates that such cognition is directed to the conditions of the possibility of a priori cognition, which includes the sensible conditions presented in the *Aesthetic* as well as the intellectual conditions discussed in the *Analytic*. It also calls to mind the question posed in the *Prolegomena* and subsequently presented as the essential concern of Kantian transcendental philosophy: how are synthetic judgments possible a priori? Although it is possible to read the first edition definition in such a way as to make it compatible with this, its reference to ‘a priori concepts of objects in general’ suggests the Wolffian conception of the transcendental as concerned with the nature of objects in general or as such, which is the subject matter of ontology or *metaphysica generalis*.⁴ Accordingly, a contemporary of Kant confronting this definition for the first time would likely construe it in ontological terms.

Indeed, this is just what was done by Kant’s earliest critics, which helps us to understand their crude misinterpretations of transcendental idealism. Thus, the notorious ‘Garve-Feder’ or ‘Göttingen Review’ characterizes transcendental idealism as a system of ‘higher idealism’, which, by encompassing both matter and spirit, ‘transforms the world and ourselves into representations’.⁵ As misguided as this may be as a reading of Kant, it seems natural enough, if one construes ‘transcendental’ in the ontological sense.

Not surprisingly, Kant was incensed by this review and responded sarcastically in a well-known note in the *Prolegomena*, where he insists that it involves a complete misrepresentation of his view. Far from referring to what is higher than or beyond experience (as claims about things or objects in general unavoidably do), Kant points out that he takes the term to refer to what make experience possible (*Prol.* 4: 373n.). Nevertheless, in an apparent recognition of the ambiguity of his initial definition, he withdraws the term ‘transcendental’ and renames his idealism ‘formal’, or preferably ‘critical’, ‘in order to distinguish it from the dogmatic idealism of Berkeley and the skeptical idealism of Descartes’ (*Prol.* 4: 375).⁶

Our second text is a well-known remark, which Kant advises the reader to keep well in mind, since its import extends to everything that follows, namely, ‘that not every a priori cognition must be called transcendental, but only that by means of which we cognize that and how certain representations (intuitions or concepts) are applied entirely a

⁴ Wolff describes ontology, which he equates with first philosophy, as ‘that part of philosophy which treats of being in general and of the general affections of being’. And he thereby defines it as ‘the science of being in general, or insofar as it is being’ (Christian Wolff, *Preliminary Discourse on Philosophy in General*, tr. Richard J. Blackwell, Indianapolis and New York: Bobbs-Merrill Co., 1983, 39). Similar formulations are to be found elsewhere in Wolff and in Baumgarten.

⁵ This citation is taken from *Kant’s Early Critics: The Empiricist Critique of the Theoretical Philosophy*, ed. and tr. Brigitte Sassen (Cambridge: Cambridge University Press, 2000), 53.

⁶ See also B518–519n., where Kant remarks that he occasionally also terms his idealism ‘formal’ to distinguish it from *material* idealism, which equates with ‘the common idealism that itself doubts or denies the existence of external things’.

priori, or are possible (i.e., the possibility of cognition or its *a priori* use)' (A56/B80–1). This is the, to us, more familiar conception of the transcendental, according to which it is distinguished from first-order *a priori* knowledge in being a second-order knowledge of the conditions of the possibility of the former. Determining these conditions and with them the limits of *a priori* knowledge is, therefore, the task of transcendental philosophy.

For present purposes, however, what is particularly noteworthy about this remark is that, immediately upon articulating this view of transcendental cognition, Kant adds parenthetically:

Likewise the use (*Gebrauch*) of space about all objects in general would also be transcendental; but if it is restricted solely to objects of the senses, then it is called empirical. The difference between the transcendental and the empirical therefore belongs only to the critique of cognitions and does not concern their relation to their object. (A56–7/B81)

Without apprising the reader of the fact, Kant here reverts to the traditional conception, according to which transcendental cognition is concerned with things or objects in general and transcendental philosophy is equated with ontology. Furthermore, the passage as a whole embodies two distinct contrasts: one is between transcendental and *a priori* cognition; the other between a transcendental and an empirical *use* of concepts.⁷ And whereas the former is a matter of level, the latter is a matter of scope. Thus, while the transcendental use of a concept (in this case space) consists in its application to objects (or things) in general, an empirical use is restricted to the subset of objects that are in some manner empirically accessible.

Moreover, these appeals to the traditional conception of the transcendental are not isolated occurrences, which might be regarded, in the manner of patchwork theorists, as signs of a reversion on Kant's part to a pre-critical standpoint.⁸ Nor are they limited to the first edition, since, unlike the latter's ambiguous stipulative definition of 'transcendental', the contrast between the transcendental and empirical use of the concept of space is contained in both editions. In fact, we shall soon see that Kant systematically uses the term in precisely this way in contrasting a transcendental with an empirical use of the categories and principles of pure understanding as well. In both cases, it amounts to a contrast between a generic use with respect to all objects and one restricted to a particular domain of objects, namely, phenomena or objects of possible

⁷ The contrasts between transcendental and *a priori* and transcendental and empirical contained in this passage are but two of several such contrasts to be found in Kant's work. Without claiming exhaustiveness, these include the transcendental-logical, transcendental-transcendent, and transcendental-metaphysical. As is often the case, the best way to understand Kant's use of a technical term is to see with what he contrasts it.

⁸ This is emphasized by Gottfried Martin, who remarks that the meaning of 'transcendental' that comes from the ontological tradition is still alive in Kant (*Kant's Metaphysics and Theory of Science*, tr. Peter Lucas, Manchester: Manchester University Press, 1955, 37). My understanding of the way in which it is alive differs markedly from Martin's, however, since, as the title of his book suggests, he tends to offer an ontological reading of the *Critique*.

experience. Indeed, in interpreting Kant, it is crucial to keep in mind that the heading of the third chapter of the *Analytic of Principles* is 'On the ground of the distinction of all objects in general into phenomena and noumena' (A235/B294), which indicates that the latter (and their correlates appearances and things in themselves) are to be understood as the two species of objects in general.

II

These reflections on the ambiguities in Kant's use of the term 'transcendental', should put us in a better position to understand his contrast between the two generic forms of transcendentalism, which he draws in two places in the *Critique*. In the first, which is located in the A-version of the Paralogisms, he characterizes transcendental realism as the view 'which regards space and time as something given in themselves (independent of our sensibility)' and suggests that such a realist 'interprets outer appearances . . . as things in themselves, which would exist independently of us and our sensibility and thus would also be outside us according to pure concepts of the understanding'. Correlatively, transcendental idealism is defined as the doctrine that all appearances are 'to be regarded as mere representations and not things in themselves, and accordingly that time and space are only sensible forms of our intuition, but not determinations given for themselves or conditions of objects as things in themselves' (A369).

Kant's second presentation of the contrast is located in his analysis of the Antinomy. Kant there accuses transcendental realism of making 'modifications of our sensibility into things subsisting in themselves, and hence makes mere representations into things in themselves'; whereas transcendental idealism affirms that 'all objects of an experience possible for us, are nothing but appearances, i.e., mere representations, which, as they are represented, as extended beings or series of alterations, have outside our thoughts no existence grounded in itself' (A490–1/B518–19).

According to both accounts, the dispute concerns primarily the nature of space and time and, as a consequence of this, the nature of things encountered in them—what Kant terms 'appearances'. In essence, Kant's charge is that by viewing space and time as 'given in themselves', that is to say, given independently of the conditions of sensibility, transcendental realism conflates spatiotemporal entities, which for the transcendental idealist are 'mere appearances', with things in themselves.

The presence of his new sense of 'transcendental' in Kant's account of his idealism is clear. The operative assumption is that space and time are forms of human sensibility, which, as such, are applicable to things as they appear, but not to these same things as they exist in themselves apart from their relation to human sensibility. To consider things in the latter manner is to consider them as objects of a putative 'pure understanding', whose thought is unconstrained by any conditions of sensibility. Inasmuch as Kant's analysis has supposedly shown that space and time are a priori conditions of cognition, which reflect the peculiar nature of human sensibility, we can appreciate

both why he initially termed his form of idealism transcendental and, in light of the ensuing misunderstanding, came to prefer the label 'critical'. Simply put, this idealism is based upon an epistemological reflection on the conditions of our cognition rather than a metaphysical reflection on the nature of its objects.

Unfortunately, Kant never explains how the term 'transcendental' is to be understood with respect to the realism which he opposes to his idealism. Rather, he seems content to equate the latter with any position that views space and time as 'given in themselves', which entails that things given in space and time ('appearances') are taken as things in themselves. But lacking any clear understanding of what it means to hold that space and time are given in themselves, it remains unclear what this amounts to and, therefore, why transcendental realism should be regarded as a form of *transcendentalism*, which, as such, might provide a valuable clue for understanding transcendental idealism.

These points are clarified, however, if we interpret transcendental realism in light of the traditional sense of 'transcendental'. First, it gives some content to the otherwise vacuous notion by suggesting that what unites its various forms is the non-trivial assumption that spatiotemporal predicates, however they be understood, are properly applicable to things in general. To be sure, this leaves room for a wide variety of positions, but it none the less suggests that they constitute a genuine family. Second, it makes it readily understandable why Kant should apply the label 'transcendental' to this family of positions and juxtapose it to his transcendental idealism.

Nevertheless, it might be objected that such a reading lacks a firm basis in the texts, since Kant does not refer to the conception of things or objects in general in either of his two accounts of transcendental realism. Instead, he refers to the more familiar conception of things in themselves and construes transcendental realism as the position that mistakenly treats mere appearances as if they were things in themselves. How, then, can my suggested reading be reconciled with what Kant actually says about transcendental realism?

The answer to this question turns on the relationship between the concepts of a thing or object in general and a thing in itself. Consider Kant's account of the putative transcendental use of the categories and their associated principles, by which he understands one with regard to the cognition of things in general.⁹ It is noteworthy that Kant contrasts this with an *empirical* use, which is restricted to the cognition of objects of possible experience or appearances, rather than with a *transcendent* one. Inasmuch as the latter does not merely ignore the boundaries of possible experience, but involves

⁹ In addition to the passages to be cited, Kant refers to a putative transcendental use of the pure concepts and/or their associated principles at A139/B178, A19/B266, A242, A246/B303, A247/B304, A296/B352–3, A402–3, A515/B544. In most of these places it is contrasted with a legitimate empirical use. In the Transcendental Deduction, however, Kant views this distinction in a quite different way with respect to the faculties of sense, imagination, and apperception, each of which is claimed to have a legitimate transcendental use as well as an empirical one (A94/B127). But here 'transcendental' refers to their function as conditions of the possibility of experience, which obviously does not involve any reference to things in general.

principles that explicitly lead the understanding to transcend them, it does not enter the picture at this point.¹⁰

Although Kant defines the categories (nominally) as concepts of an object in general, he denies that they have a transcendental employment, since they would then lack the schemata necessary to link them up to the world.¹¹ In fact, it is precisely because it involves an illicit transcendental employment of the understanding that Kant proclaims famously that: '[T]he proud name of ontology, which presumes to offer synthetic *a priori* cognitions of things in general in a systematic doctrine . . . must give way to the modest one of a mere analytic of the pure understanding' (A247/B304). As this passage indicates, the crux of the problem lies in the pretension of ontology to provide synthetic *a priori* cognitions of things in general, when all that it can actually deliver are analytic judgments, which unpack what is already thought or presupposed in the concept of an object as such.

At the same time, however, Kant also suggests that if (*per impossibile*) the categories had a transcendental employment, 'as conditions of the possibility of things in general, they [could] be extended to objects in themselves (without any restriction to our sensibility)' (A129/B178). And, in a similar vein, he states that, 'The transcendental use of a concept in any sort of principle consists in its being related to things *in general* and *in themselves*' (A238/B298). These, as well as other passages, make it evident that Kant considered the application of a concept to things in themselves to be a direct consequence of its application to things in general.¹²

Since we have seen that the concept of things or objects in general is taken over from the traditional ontology, whereas the concept of a thing in itself is a product of Kant's transcendental critique, where it is contrasted with the concept of an appearance, it is precisely at this point that the two conceptions of the transcendental come together. Moreover, both are required for an understanding of Kant's contrast between the two forms of transcendentalism and the turn from one to the other.

The key lies in the different ways in which the concept of things in general and in themselves involve the thought of an independence from sensibility and its putative conditions. In the former case, it is simply a matter of ignoring or abstracting from any relation to human (or any other kind of) sensibility. In reflecting on what pertains to the concept of a thing as such, or *qua* thing, one is not interested in what may pertain

¹⁰ See A296/B352–3, where Kant denies that 'transcendental' and 'transcendent' are the same.

¹¹ Although Kant explicitly denies that the pure categories can be defined at A245, he there also states that they are 'nothing other than the representations of things in general, insofar as the manifold of their intuition must be thought through one or another of these logical functions'. For Kant's definitions, see A93/B126, A248/B305, A253, and A290/B346. Thus, following Lewis White Beck, I think it best to regard Kant as providing a nominal definition of the categories. See Beck, 'Kant's Theory of Definition', *Studies in the Philosophy of Kant* (Indianapolis: Bobbs-Merrill Co., 1965), 61–73.

¹² See e.g. Bxxvii, where Kant claims that, if the distinction between things as objects of experience and the very same things as things in themselves were not drawn, then the principle of causality would be valid of things in general as efficient causes; and B410, where Kant suggests that if the rational psychologists were right, synthetic propositions 'could reach as far as things in general and in themselves'.

to things only in virtue of any contingent relationship in which they might stand to human sensibility. As the concept of an object thought through a pure understanding, however, the concept of a thing as it is in itself refers to things as they are thought independently of their relation to sensibility. Accordingly, this concept involves an explicit denial rather than merely an abstraction from any relation to sensibility.

Nevertheless, this difference does not prevent Kant from moving seamlessly from the former conception to the latter; for since whatever is predicated of things in general is predicated of them absolutely or in every relation, it must also be predicated of them as they are in themselves.¹³ Consequently, there is no incompatibility between viewing transcendental realism as attributing spatiotemporal predicates to things in general and Kant's charge that such realism is guilty of attributing them to things in themselves. On the contrary, this helps us to understand why, when seen from Kant's point of view, the transcendental realist would invariably make the latter mistake. It is precisely because transcendental realism inflates spatiotemporal predicates into predicates of things in general that it unavoidably attributes them to things in themselves as well. After all, what is true of the genus is also true of the species.

We can also see from this that transcendental idealism is transcendental in *both* senses of the term distinguished earlier. First, we have seen that it is transcendental in the distinctively Kantian sense because the restriction results from a transcendental reflection on the condition of cognition. Second, it is transcendental because, as result of this reflection, it denies transcendental status (as traditionally understood) to spatiotemporal predicates. As such, it presupposes the ontological conception and defines itself in opposition to it. It is because of this that I believe it fair to characterize Kant's transcendental turn as a turn from one form of transcendentalism to another.

If this analysis is correct, it further suggests that transcendental idealism may be understood as the position which refuses to inflate these predicates in this way, but instead limits their application to objects of possible experience. Indeed, we shall see that, so construed, transcendental idealism just is empirical realism, at least insofar as the latter is defined in opposition to transcendental realism, as consisting in the limitation of the scope of such predicates to objects of possible experience. What transcendental idealism adds to this is merely an account of the justifying grounds for this limitation, namely, the source of these predicates in human sensibility. Demonstrating this is the task of the Transcendental Aesthetic to which we now turn.

III

After a series of definitions, the Transcendental Aesthetic begins in a seemingly ontological mode by raising the question: what are space and time? Four possibilities are proposed. They might be: (a) actual entities (substances), (b) determinations

¹³ See A324–5/B380–2.

of things (accidents), (c) relations of things that 'would pertain to them even if they were not intuited', or (d) 'relations that only attach to the form of intuition alone, and thus to the subjective constitution of our mind, without which these predicates could not be attached to anything at all' (A23/B37–8).¹⁴ Although Kant does not call attention to the fact, the first three represent the traditional ontological options and, therefore, apply to things in general. Of these, the second and third correspond respectively to the Newtonian and Leibnizian views, which were the main competitors at Kant's time.¹⁵ The fourth is Kant's 'critical' view and reflects his rejection of the whole ontological framework in which the question had traditionally been posed, one in which it is assumed that *whatever* status is given to spatiotemporal predicates they apply with strict universality, that is, to things or objects in general. Thus, from Kant's point of view, the dispute between the Newtonians and the Leibnizians is reduced to a family quarrel amongst transcendental realists.

Seen in this light, Kant's basic charge against transcendental realism is that it erroneously assumes that spatiotemporal predicates have an unrestricted scope. Against this, Kant claims not merely that such predicates are relational, a point on which he agrees with Leibniz, but also, and more significantly, that they are attached to things only in virtue of their relation to our forms of intuition, that is to say, they are predicable only of things *qua* sensibly intuited. Given Kant's transcendental distinction, it follows from this that they are not predicable of things as they are in themselves and, therefore, not of things in general. The principle here is the obverse of the one just affirmed: what is *not* predicable of a species is *not* predicable of the genus.

¹⁴ That Kant lists four possibilities, rather than merely the three that I suggested in the 1st edn of *Kant's Transcendental Idealism*, has been noted by Lorne Falkenstein, *Kant's Intuitionism: A Commentary on the Transcendental Aesthetic* (Toronto: University of Toronto Press, 1995), 147. As he correctly notes, this was already pointed out by both Hans Vaihinger, *Commentar zu Kants Kritik der reinen Vernunft*, ii (Stuttgart: W. Spemann, 1881–92), 131–4, and Martin, *Kant's Metaphysics and Theory of Science*, 11–12. Moreover, there are several other texts in which Kant clearly distinguishes between these possibilities, including ID 2: 400 and 403, *Reflexion*, 5298, 18: 146–147, and *Reflexion*, 5404, 18: 174. Nevertheless, at least from the time of the Dissertation, Kant effectively assumed that the only two alternatives to his own view worthy of serious consideration were the Newtonian and the Leibnizian positions. Thus, in effect, he recognized only three 'live options'.

¹⁵ It might be wondered why Kant should claim that the Leibnizians 'ontologized' space and time, since, like Kant, Leibniz held that they were 'ideal' in the sense that they pertain only to phenomena. Quite apart from the question of the adequacy of his interpretation of Leibniz, however, it is clear that Kant's Leibniz is committed to the thesis that spatiotemporal predicates (properly construed) are applicable to things in general. As Kant saw it, this is because 'Leibniz intellectualized the appearances' (A271/B327), by which Kant meant that for Leibniz the difference between what 'appears' or is sensibly represented and what is grasped intellectually or conceptually is a matter of degree of clarity and distinctness rather than of kind. Thus, while Kant was well aware that Leibnizian monads are not in space and time, he also insisted that for Leibniz the spatiotemporal relations holding between the *phenomena bene fundata* are reducible in principle (though not for us) to the purely conceptual relations supposedly holding at the monadological level. Moreover, the latter relations are clearly thought by the Leibnizians to apply to things in general. I discuss this issue in *Kant's Transcendental Idealism* (2004), 29–31, and at greater length in 'Kant and the Two Dogmas of Rationalism', in Alan Nelson (ed.), *A Companion to Rationalism* (Oxford: Blackwell, 2005), 343–59.

Kant's well-known formula for this is that space and time are empirically real but transcendently ideal. Since I cannot here consider the complex and deeply controversial line of argument behind this claim, I shall limit myself to the more modest, but non-trivial task of attempting to articulate what it means.¹⁶ To begin with the obvious, to claim that space and time are transcendently ideal is to deny that they are transcendently real, which is logically equivalent to denying that spatial and temporal properties and relations are predicable of things in general. Of itself, however, this claim is compatible with their not being truly predicable of anything at all. This is ruled out by the thesis that space and time are empirically real.

What, then, does it mean to claim that space and time are empirically real? On the one hand, it is to deny that they are empirically *ideal*, which would entail that, like secondary qualities, spatial and temporal properties and relations are not truly predicable of empirical objects. Positively expressed, this means that their properties and relations *are* truly predicable of objects of possible experience. It is also, however, to deny that they are *transcendently* real, which, once again, means that these properties and relations are not predicable of things in general. Accordingly, to claim that space and time are empirically real is to claim that they are *merely* empirically real, which is a consequence of their being transcendently ideal. Combining these points, we arrive at the result that space and time are 'really real' in the sense that their properties and relations are truly predicable of a domain of objects, but that in virtue of their source in the nature of our sensibility, this domain is restricted to objects of possible experience. Moreover, it follows from this that transcendental idealism, *even as it appears in the Aesthetic*, is best seen as a deflationary proposal rather than as an ontological thesis in direct competition with the various forms of transcendental realism.¹⁷

Before leaving the Aesthetic, I would like to pursue the last point a bit further, since it is deeply controversial. Recall that the first three alternatives that Kant entertains exhaust the recognized ontological options. This naturally suggests the question of how we are to understand the fourth alternative for which Kant opts. There seems to be two possibilities: we can regard it either as a novel move *within* ontology or as a radical alternative *to* ontology. The ontological framework in which Kant poses the issue suggests the former, and this is also the way in which the Aesthetic is usually read. Nevertheless, I believe it to be not only possible but preferable to take it in the latter

¹⁶ I analyse and attempt to defend Kant's argument in *Kant's Transcendental Idealism* (2004), esp. pp. 122–32. Here, I wish merely to point out that a direct and important corollary of this reading is the assignment of a central place to the intuition arguments of the Aesthetic, since it is through these alone that Kant attempts to link the representations of space and time with human sensibility. Assuming their apriority, if space and time were assigned to the understanding rather than to sensibility, Kant would have had to conclude that they are predicable of things in general. Accordingly, it is of no little significance to the 'critical project' for Kant to be able to show that: 'Space is not a discursive or, as is said, general concept of relations of things in general, but a pure intuition' (A24–5/B39).

¹⁷ I emphasize the Aesthetic because it is here that the ontological reading of Kant's ideality thesis seems most compelling. Thus, if this reading can be challenged here, the stage is set for a comprehensive non-metaphysical interpretation of transcendental idealism.

sense. In Kant's own terms, the alternative is to consider space and time as 'two sources of cognition' (A38/B55), that is, as conditions of our *cognition* of things, rather than as themselves either things (substances), properties, or relations of things as such.

Admittedly, it remains possible to dig in one's heels and insist that Kant's claim still must be understood ontologically, since it involves at least a negative ontological thesis, namely, that spatial and temporal properties and relations are *not predicable* of things in general because they are *not predicable* of things as they are in themselves. In my view, however, this significantly downplays the critical thrust of Kant's position and, as a result, leads to considerable misunderstanding. In order to preserve this thrust, we must recall and apply to the Aesthetic the lesson about the need to give up the 'proud name of ontology' that was taught in the Analytic. In fact, this lesson is applicable to the categories in the Analytic only because it was already applied to space and time in the Aesthetic. Since according to Kant the reach of the categories extends as far as their sensible conditions or schemata, it is precisely by limiting sensibility that the categories are kept within their legitimate sphere of operation. In other words, if space and time were applicable to things in general, then so would the categories be, from which it would immediately follow that they are likewise applicable to things in themselves. Thus, when at the end of the Analytic Kant introduces the concept of the noumenon and assigns it the task of curbing the 'pretension of sensibility' (A255/B311), he is presupposing the results of the Aesthetic where spatiotemporal predicates are limited to objects qua sensibly represented.

IV

Whereas in the Aesthetic Kant argues directly for transcendental idealism by attempting to show that the spatiotemporal structure of human experience is grounded in the nature of our sensibility, which precludes its projection onto things in themselves and, therefore, things in general, in the Antinomy of Pure Reason he argues indirectly for the same idealism by endeavoring to demonstrate that this projection, which is definitional of transcendental realism, leads to a set of contradictions that disappear under the transcendently idealistic restriction of the scope of this structure to the domain of possible experience. I shall here approach the latter thesis by considering why transcendental realism unavoidably falls into these contradictions, how transcendental idealism supposedly avoids them, and what this tells us about the nature of transcendental idealism.

Confining ourselves to the first antinomy, the basic problem is that the endeavor to determine the age and size of the spatiotemporal world leads to two equally warranted but incompatible conclusions, namely, that this world is infinite in duration and extent and that it is finite in both respects. Moreover, this suggests that the very concept of a spatiotemporal world, like that of a square circle, must be self-contradictory, since it generates contradictory entailments.¹⁸

¹⁸ See *Prol.* 4: 341, where Kant poses the issue in this logical form. It is also noteworthy that Arthur Collier, with whose work Kant was probably familiar, used virtually the same antinomial argument in an attempt to

In essence, Kant's resolution of this antinomial conflict consists in claiming that these alternatives are mere contraries rather than genuine contradictories and that both are false. This is because these conclusions share the unwarranted assumption that the spatiotemporal world is a 'whole existing in itself' (A506/B534).¹⁹ Given this assumption, it follows that the world must have a determinate duration and extent; that is, to say, there must be some 'fact of the matter' regarding the age and size of the world. But once this assumption is denied the contradiction disappears.

Kant also recognizes, however, that this assumption is not easily denied; for even if we are unable to determine whether the world is finite or infinite in the relevant respects, it appears evident that it must be one or the other. Kant's explanation for this turns on his doctrine of transcendental illusion, which I can here sketch only in the broadest terms.²⁰ The locus of this illusion is the seemingly innocent principle of reason: 'If the conditioned is given, then the whole series of all conditions for it is also given' (A497/B525). Since the conditions at issue here are just the components of the sensible world, to assume that the whole series of such conditions is given is to assume that this world constitutes a 'whole existing in itself'. What makes this principle seemingly innocent is that it merely expresses the explanatory requirement to seek the conditions for every conditioned and not to stop until the absolute totality of these conditions is attained. What makes it illusory is that such an absolute totality can never be given (whether as finite or infinite) in a possible experience. But since the cosmological dispute concerns the duration and extent of the spatiotemporal world, it purports to be about something that is, at least in principle, an object of possible experience.

It is at this point that transcendental realism enters the story. Basically, the transcendental realist is led by this illusory principle to 'hypostatize' the spatiotemporal world, that is, to treat it as a higher order empirical object, even though such an object could never be given in accordance with the conditions of possible experience. In view of our previous analysis, this hypostatization can be seen as the result of the inflation of an

prove that 'an external world, whose extension is absolute, that is, not relatively depending on any faculty of perception', is self-contradictory (*Clavis Universalis*, in *Metaphysical Tracts by English Philosophers of the Eighteenth Century*, ed. Samuel Parr, London: Edward Lumley, 1837, 46–50).

¹⁹ Although this assumption does not enter as a premise into either the thesis or antithesis argument of any of the antinomies, it underlies the cosmological debate as a whole. In particular, it makes it possible for each party to argue apagogically from the falsity of the alternative to the truth of its own claim. This also enables Kant to vouch for the soundness of each of the proofs, while at the same time claiming that the whole dispute is based on a misunderstanding. Admittedly, these proofs remain highly controversial, but I have endeavored to defend those of the first and third antinomies against the standard objections in *Kant's Transcendental Idealism* (2004), 366–84.

²⁰ For the definitive account of transcendental illusion, see Michelle Grier, *Kant's Doctrine of Transcendental Illusion* (Cambridge: Cambridge University Press, 2001). My own systematic discussion of the topic, which is greatly indebted to Grier's but differs on some points, is to be found in *Kant's Transcendental Idealism* (2004), 322–32.

empirical realism regarding objects of possible experience into a transcendental one encompassing things in general. As before, if spatiotemporal predicates are thought to attach to things in general, then any subjective conditions of our experience of them are irrelevant to their nature and status.

Although I believe that this concurs with Kant's own account, he poses the issue in somewhat different terms. According to Kant, the key to the problem lies in the assumed 'givenness' of the conditions. Whereas a transcendental realist naturally assumes that the totality of conditions is actually 'given' (*gegeben*) together with any conditioned (in the timeless manner in which the totality of its premises is 'given' together with the conclusion of a bit of deductive reasoning), Kant points out that, as spatiotemporal entities or occurrences, these conditions are merely 'given as a problem' (*aufgegeben*) (A498/B526), by which he means that they are accessible only through the regress or 'synthesis' connecting something conditioned with its conditions. Thus, within this context, it no longer makes sense to speak of an absolute totality of conditions, save perhaps as a regulative idea to be approached asymptotically but never attained.

For Kant, however, the transcendental realist is blocked from appealing to the *gegeben*–*aufgegeben* distinction because he regards the synthesis connecting something conditioned with its conditions as one 'of the mere understanding, which represents things *as they are* without paying attention to whether and how we might achieve acquaintance with them' (A498/B527).²¹ While the expression 'things *as they are*' suggests the thought of things as they are in themselves, a synthesis of 'the mere understanding', which involves the pure categories, may just as well be taken as concerned with the thought of objects in general. In either event, the point is that it is precisely because the transcendental realist systematically ignores the manner in which these synthesized conditions are cognitively accessible in a possible experience that he tacitly assumes that they must be given in their totality in the way in which a complete set of premises is given together with a conclusion. Moreover, it is because he assumes this to be the case that such a realist is unable to resist the pull of the illusory principle that the absolute totality of conditions is given, even though it is not accessible to the human mind. Finally, it is because of this inability that the transcendental realist is unavoidably led to assume that the world is a 'whole existing in itself', such that it must be either finite or infinite in the relevant respects.

Although it remains natural, even for the transcendental idealist, to *think* that there must be a totality of conditions, that is, some ultimate fact of the matter

²¹ Since by such a synthesis Kant understands one that makes use merely of the pure or unschematized categories, in stating that it represents things as they are he is clearly not suggesting that it provides cognition of things as they are in themselves. His point is rather that it regards the items synthesized (the conditioned and its conditions) as a collection of objects whose nature is fixed apart from any sensible conditions that may be necessary for us to access them, i.e. as a collection of things considered as existing in themselves. It should also be kept in mind that one of the ways in which Kant characterizes the thought of things as they are in themselves is as objects of a 'pure understanding'.

about the duration and extent of the spatiotemporal world, such an idealist is also aware that this thought is illusory and is therefore able to avoid succumbing to it.²² What makes this avoidance possible is the realization that the scope of the principles underlying the synthesis connecting a conditioned with its conditions, like the scope of the spatiotemporal conditions of experience on which they depend, is limited to possible experience. In other words, transcendental idealism functions as a defense (indeed the only defense) against transcendental illusion by means of its restriction of the scope of spatiotemporal predicates to objects of possible experience.

This puts us in a position to deal succinctly with Kant's much maligned 'indirect proof' of transcendental idealism, which he casts in the form of a dilemma:

If the world is a whole existing in itself, then it is either finite or infinite. Now the first as well as the second alternative is false (according to the proof offered above for the antithesis on the one side and the thesis on the other). Thus it is also false that the world (the sum total of all appearances) is a whole existing in itself. From which it follows that that appearances in general are nothing outside our representations, which is just what we mean by their transcendental ideality. (A506–7/B534–5)

Even though transcendental realism is not mentioned in this argument, it seems evident that it is Kant's target, since such realism is committed to the view that the world is a whole existing in itself, which is the source of the contradiction. Accordingly, the argument proceeds by denying transcendental realism in virtue of this commitment, and this denial is taken to entail transcendental idealism. Clearly, an implicit premise of this argument is that these two forms of transcendentalism, unlike the theses and antitheses of the various antinomies, are genuine contradictories. Otherwise the affirmation of transcendental idealism through the negation of transcendental realism would be a non sequitur of more than numbing grossness.

Although this raises many questions, one I wish to pose by way of a conclusion, is seldom asked, namely, what must be the nature of transcendental idealism, if its truth were to be established by something like the line of argument shown here? The reason it is seldom asked is that it is usually assumed that transcendental idealism is a subjectivistic metaphysical doctrine. Indeed, this is suggested by Kant's own account of transcendental ideality as the thesis that 'appearances in general are nothing outside our representation'. And since it does not appear that any such metaphysical conclusion is warranted by Kant's argument, the latter is almost universally deemed a failure.

²² The essential point here, which has been developed at length by Grier (*Kant's Doctrine of Transcendental Illusion*) is the distinction between the illusion and the metaphysical fallacies it generates (which include those committed by the participants in the antinomial conflict). Whereas the former is unavoidable, the latter are not, even though they are based on this illusion. As Grier shows, failure to keep this distinction in mind underlies much of the confusion regarding Kant's critique of metaphysics in the *Dialectic*.

Things look rather different, however, if transcendental idealism is understood as a form of critical transcendentalism, which, in contrast to transcendental realism, insists upon the restriction of spatiotemporal predicates to the domain of possible experience. For we have seen that to be a transcendental idealist in this sense just is to be an empirical realist. And given the difficulties to which transcendental realism is subject, the real question becomes why would one want to be anything more.

2

On Reconciling the Transcendental Turn with Kant's Idealism

Karl Ameriks

1. Looking Beyond the Turn

Ever since P. F. Strawson's extremely influential account in the *The Bounds of Sense*, there has been a widespread tendency to presume that the only way to save the 'critical insight' and 'truth' of the *Critique of Pure Reason* and its revolutionary 'Transcendental Turn' is to 'withdraw from the residual metaphysical commitments' in Kant's own version of transcendental idealism.¹ Kant's critical insight is supposed to consist in a move away from metaphysics and toward a 'Copernican' procedure of justifying the universal and necessary structures needed within our experience.² These structures are called 'transcendental' simply to highlight the fact that they enable a priori knowledge and thus are free from the limitations of merely particular and contingent empirical grounds, while the path to their justification is characterized as a 'turn' precisely to highlight the fact that the structures are meant to be revealed through a purely critical technique, that is, one which asserts only what is immanently demanded by our own epistemic practices in general and withdraws from any reference to grounds in the transcendent—and ultimately opaque to us—entities of traditional dogmatic metaphysics.

The doctrine of transcendental idealism, however, supposedly infects Kant's system with the untenable paired notions of a wholly transcendent realm of things in themselves in general and a privileged ego 'in itself'; a supersensible entity whose foundational operations can be explored only through the mysterious enterprise of

¹ Sacks 2000, 215. Cf. Strawson 1966 and Bird 2006.

² I take 'experience' for Kant to mean putative *determinate theoretical cognition* of what is sensible and objective. On rare occasions he uses the term without these restrictions, e.g. in speaking of the 'experience' of the fact of morality that reveals our freedom. See Kant 1790a (*Critique of the Power of Judgment*), §91 (5: 468).

'transcendental psychology'.³ The often reiterated worry here is that, as long as any metaphysical notion of things in themselves remains in a Kantian philosophy, the hidden structures and powers of such things must undercut the guiding intention of the Transcendental Turn itself and 'demote' the genuine objectivity of whatever structures of our experience his own Transcendental Turn claims to reveal.⁴ Furthermore, as long as the metaphysical reading of Kant is taken to have to respond to this worry by saying that these structures are backed up by a 'transcendental psychology' that supposedly explains them by reference to mysterious bootstrapping or skyhooking 'acts' of a disembodied ego, this would seem to lead only to a further destabilization of whatever claims the Transcendental Turn makes.

Although these kinds of considerations may help explain why so many contemporary philosophers still broadly sympathetic to Kant have been oriented toward finding ways to develop a thoroughly *nonmetaphysical* reading of the *Critique*, there remains a quite different and neglected strategy that remains worth exploring. This strategy, which defines what I will call the Moderate Interpretation of Kant's idealism,⁵ highlights the possibility of finding a way to appreciate the concerns that lie behind the nonmetaphysical reading while still not going so far as to insist on the *antimetaphysical* view that these concerns make it necessary to 'withdraw from' Kant's 'residual metaphysical commitments' altogether. One can pursue this possibility while still agreeing that it is very important, for exegetical as well as systematic reasons, not to tie the mere notion of a Transcendental Turn too closely to the mere notion of transcendental idealism. The Turn's definitive claim, which is that there *are* immanently determinable necessary structures of our experience and its objects, is by itself a nonmetaphysical point (in the way that terms are being used here), and it therefore hardly seems identical with transcendental idealism's definitive metaphysical claim, which is that such structures should, in the end, be characterized as conditions of what is *only* our experience and its objects.

As the *Critique* develops, there are, of course, several close connections that are made between these two claims, but at first sight they certainly appear to be oriented toward distinct contexts. The Turn's claim about the existence of necessary transcendental truths, especially as it is developed in Kant's main arguments (the transcendental expositions of space and time, the transcendental deduction of the categories, the arguments for the specific principles of the analogies of experience), is a claim that is expressed in terms of epistemological issues; specific kinds of familiar cognitive claims, in mathematics as well as everyday life (concerning, for example, houses, boats, and 'determining oneself in time'), are argued to be unsustainable without accepting more basic instances of underlying a priori knowledge about the world's

³ Sacks 2000, 303.

⁴ The term 'demote' is discussed in section 5, and is taken from Wood 2005, 65. Similar worries are expressed by McDowell 1994, 44, who uses the term 'fraudulent'; and Stroud 1999.

⁵ In preferring to use the term 'moderate' here (rather than 'modest' or 'humility'), because it signifies 'in between extremes', I have been influenced by Allais 2007.

structures. Kant's transcendental idealist claim, in contrast, is presented as a metaphysical answer to a question about the ultimate ontological nature of the items that are said to be known through the a priori principles taken to be established by the Transcendental Turn; it is a claim that these items cannot be regarded as things in themselves, or their 'determinations' or 'relations', because all these items are essentially determined by space and time, and space and time have specific structural features that keep them from being able to characterize any feature of things in themselves as such.⁶ The Moderate Interpretation builds on this basic distinction between the definitions of the Turn and of Kant's idealism by adopting the neglected strategy of seeking an at least understandable (even if not systematically compelling for all contemporary philosophers) way of holding on to the Turn even while interpreting the *Critique* as also containing—as Kant certainly appears to have intended—a weighty metaphysics of things in themselves that is still idealistic in a relevant but not evidently absurd sense.

The success of such a strategy obviously depends on the Moderate Interpretation's presentation of a new, plausible, and non-threatening reading of key terms such as 'in itself' and 'transcendental idealism'. Before such a reading is developed, the seriousness of the challenges that it faces should be openly acknowledged. What especially complicates matters here is the fact that, right after carrying out key instances of the Turn and arguing that certain necessary structures, such as pure intuitions and forms of judgment, exist in our knowledge, Kant himself repeatedly moves on very quickly to an extra point by adding the assertion that his idealism alone can 'make intelligible' the necessary status of the knowledge of the principles that have just been critically established.⁷ This specific assertion about 'making intelligible' is by no means easy to understand or defend, and it has been heavily disputed throughout the history of Kant scholarship. Whatever its ultimate status, the most important thing to note at this point about the assertion is just that it amounts to a distinct claim. It is not by itself identical with the defining claims of either the Turn or idealism, and since it is obviously more complex and hence less fundamental than these claims, it might be bracketed, qualified, or perhaps even jettisoned without eliminating the others. The assertion does, nonetheless, understandably give rise to the main worries that concern interpreters who raise the issue of Kant's 'transcendental psychology'. As soon as some kind of hidden nonempirical 'invariance' of the mind seems to be invoked to 'explain' the non-invariance of our knowledge, the fear can develop that this invisible side of the mind could be tricking us into holding to principles that are limited and relative in a way that undercuts their legitimacy.⁸ Note, however, that this fear concerns only an alleged possibility—a worry that might be countered by further investigation—and it is significant that Kant's own discussion at this point does not actually develop a positive 'psychological' explanation of specific knowledge claims (or use the term 'transcendental

⁶ Kant 1781/1787, A23/B37 (*Critique of Pure Reason*).

⁷ B41. See also Kant 1783 (*Prolegomena*) §9 (4: 482).

⁸ Sacks 2000, 304.

psychology') but seems most concerned with ruling out what are taken to be dogmatic and wholly unhelpful alternatives.

Once one sees that Kant's main concern here is to counter any proposed 'explanation' of our a priori knowledge that (unlike his own idealist and more restricted position) amounts to a claim of a demonstrable 'necessary agreement' of this knowledge, and all that it conditions, with objective structures that are simultaneously taken to be structures of *things in themselves* (and not merely with objects of experience), one can see—without getting into any 'psychology'—why he would emphasize that it is unclear how any unrestricted and non-idealistic claim like this (or even a totally agnostic position) would really have a chance of making matters 'intelligible'—especially since here Kant also believes that he has several specific arguments against each one of the main traditional metaphysical ways (e.g. in Descartes, Locke, Berkeley, Hume, Leibniz, and Newton) of conceiving of space and time as things in themselves. In addition, to insist that Kant's idealism must stand or fall with the status of transcendental psychology, one would also need to think that the explanatory power of some kind of psychology—especially to account for geometry—is *required* just to make any *sense* of Kant's doctrine of the transcendental ideality of space and time. This thought conflicts, however, with the clear historical and systematic fact that Kant's doctrine has a number of other grounds, including especially the distinct arguments of the First Antinomy, which make no reference to 'explaining' geometry.

Here it is also important, for general systematic purposes, to keep in mind that there are a number of ways, quite apart from any arguments of the *Critique*, in which one could remain a kind of non-subjectivist and still hold to an innocent meaning for the mere idea that space and time, or even human knowledge in general, should not be claimed to characterize things in themselves. There are numerous nonskeptical and understandable ways in which other reputable scientists and philosophers can be taken to have in fact held such a position—ways that can involve peculiarities in specific notions of space and time, or of the 'in itself', and that are not dependent on (although they might be consistent with) whatever Kant means by considerations that would positively 'make intelligible' our mathematical principles, or on anything that is basically a matter of psychology.

Somewhat similar points can be made in response to a related objection that the mere metaphysical notion of a thing in itself conflicts with the Transcendental Turn's conclusion that 'individuation . . . is internal to the context of experience', that is, that the only means we have for *warranting* the theoretical determinations that distinguish objective items from one another and characterize their identity is to rely at some point on matters such as familiar spatiotemporal paths, in contrast to any alleged *insight* into how the detailed ontology of things in themselves may go.⁹ This objection is not fatal, however, because Kant himself clearly emphasizes precisely this conclusion of

⁹ Sacks 2000, 302.

the Turn—and he can and does do so without at all rejecting as totally impossible and nonsensical the *very idea* of some kind of principles holding for the individuation of objects as things in themselves that might not be the same as the principles we use for the objects of experience.¹⁰ This is admittedly an unusual idea, but whatever those unknown principles might be, an extra argument would have to be made to show that the mere idea of them must *undercut* any of the particular distinctions that we make within our experience.

2. Idealism in Context

The antimetaphysical reading of Kant that has just been discussed is best understood as part of a general pattern of reactions to Kant's transcendental idealism, a dominant pattern that has for a long time involved an oscillation between two related and extreme interpretive approaches. At one extreme there is the neo-Strawsonian approach, which stresses the *Critique's* insistence on empirical realism and minimizes the issue of idealism by dismissing—either as meaningless, inconsistent, totally unfortunate, irrelevant, or not even really intended by Kant—all of the text's metaphysical talk about things in themselves and nonempirical 'affection' in perception or noumenal causality in general. At the other extreme there is the long-standing neo-Berkeleyan approach, which stresses the *Critique's* idealism but understands it in terms of a psychological and subjectivist position that leaves no room for a robust form of empirical realism, let alone a consistent affirmation of even the possibility of any kind of independent nonempirical reality. This is an old issue—already by 1787 F. H. Jacobi and others had become very influential in characterizing and criticizing transcendental idealism in similar extreme terms, even though Kant wrote his *Prolegomena* (1783) and revised his *Critique* in its B edition (1787) largely to explain how his system is not vulnerable to objections of this kind concerning either subjectivism or things in themselves.¹¹

Despite Kant's own extensive response, criticisms along Jacobi's lines continue to dominate contemporary interpretations, including very recent work by some of the best specialists. Allen Wood, for example, has come to conclude that Kant's Critical

¹⁰ See esp. A260/B316ff ('Amphiboly of Concepts of Reflection'), and Kant 1790b (*Contra Eberhard*), 8: 209n. I use the terms 'thing in itself' and 'things in themselves' interchangeably. Kant does not give an argument for a plurality of such things and would reject as dogmatic any theoretical claim to know how to individuate them, but he always starts from a core commonsense belief that it is absurd to accept monism and to deny that there is any ultimate ontological distinction between 'us' and at least some other reality.

¹¹ Kant's first edition formulations are admittedly in need of the later explanations and revisions that he made, but there are ways to defend even the perplexing earlier formulations. See Ameriks 1982/2000a, 111–14, and also the discussion of Kant and Jacobi, in Ameriks 2006, 67–88, 120, and 154–5. Like several other interpreters, Jacobi imagines that, if Kant's starting point is not a level of mere sensations but a premise about objective judgments, then 'the absolute universality and necessity of these concepts [of causality in the world] must *equally be given as a prior presupposition*' (Jacobi 1787, 336). With this move, in the phrase that I italicize, he undercuts the possibility of properly understanding the Transcendental Turn, and not only Kant's idealism.

texts must be read as repeatedly wavering between two extreme and incompatible positions on idealism.¹² The first interpretive option, which he calls the Causality Interpretation, takes transcendental idealism to be an incoherent combination of subjectivism and transcendent metaphysics, an unattractive quasi-Berkeleyan version of psychological idealism that is only made worse and inconsistent with Kant's addition of occasional, and allegedly nonsensical, claims about non-empirical affection. The only alternative to this reading that Wood examines is a non-subjectivist and non-metaphysical quasi-Strawsonian position, called the Identity Interpretation, which in effect reduces Kant's idealism to a condensed expression of the Transcendental Turn, a reminder of the fact that we are limited sensible beings, with a priori forms such as space and time playing a necessary epistemic role throughout our objective experience. Wood's systematic preferences lead him toward strongly preferring the second position, and he links it favorably to interpretations by Henry E. Allison, whose antimetaphysical reading of Kant's idealism has been extremely influential but also repeatedly challenged of late, despite the many virtues of other aspects of Allison's well-known analyses of the Critical philosophy.¹³

All this puts into striking relief the mere idea of the Moderate Interpretation, which remains as a distinctive alternative to all these extreme approaches. It alone holds out for the possibility that Kant has a consistent and nonabsurd way of maintaining a weighty empirical realism—along with whatever pure structures the Transcendental Turn uncovers—even without also denying commitment to a weighty notion of things in themselves. The remainder of this chapter will therefore be devoted to characterizing the Moderation Interpretation in more detail and to explaining how it can help in overcoming problems afflicting the extreme options that dominate Wood's representative discussion.

3. Moderation from the Beginning

The first step in seeing how Kant's system can be understood in moderate terms is to recall that the *Critique* can be read from the *very beginning* as working at a level that already accepts the core objective commitments of everyday perceptual realism, rather than as restricting itself to a starting point that is idealist in any subjectivistic sense. All too many readers begin by projecting a kind of Cartesianism¹⁴ onto the *Critique's* starting point, as if its deductions and initial discussion of representation, appearances, and perception all must already involve using, in the manner of earlier modern philosophies, an *argumentative basis* of merely subjective psychological items. In fact, however, although Kant allows the existence of a lowest psychological level of mere

¹² Wood admits that one must 'readily concede' (2005, 71) that this is a distressing hermeneutical position.

¹³ See e.g. the helpful review in Gardner 2005.

¹⁴ I say 'so-called' or use quote marks to indicate that I am not assuming that Descartes' own position can be identified with what contemporary philosophers have come to criticize as 'Cartesianism'.

sensations or feelings, which have an incontrovertible but merely subjective and indeterminate reality, he is explicit from the start that he understands the particular elements of this level, and all their associations, as noncognitive,¹⁵ and therefore as playing no role as a basis in the specific arguments of the Transcendental Turn, which aim at establishing pure and necessary structures for items already taken to be determinably objective in some way. In this way the *Critique of Pure Reason*, like Kant's other *Critiques*, can be said to begin its arguments not from the premise of a strictly individualist and subjectivist psychological perspective, but from commonly presumed objective and public claims of some kind—in this case the minimal commonsense position that we human beings are basically alike in being 'receptive' to 'objects' that are 'given' to 'us', and that are even presumed to be given to all of us in the same forms of space and time.¹⁶

From this perspective, readers actually following the *Critique's* own distinctive and non-Cartesian argumentative path can always already be beyond assuming any kind of 'fraudulent' subjectivism, and the burden of proof falls on others who would go so far as to insist that Kant is starting elsewhere, or that for some other reason he must be regarded as caught in a wholly subjectivist position incompatible with his initial statements. Similarly, when Kant goes on to affirm, as he repeatedly does, that there are items that are in some sense not merely empirically objective but also metaphysical—existing 'in itself' and as 'transcendentally real', with features transcending all that we could determine by theoretical reason—the natural reaction of charitable readers can be to proceed by trying to understand the text not as suddenly forcing the introduction of an incoherent and/or basically subjectivist theory but as merely pointing to yet another level, albeit a controversial one, of what remains a consistent and basically objective position.

4. Moderation to the End

The guiding antisubjectivist hypothesis of the Moderate Interpretation lies in its basic proposal that, just as the affirmation of the psychological realities of a first level of indeterminate, private, and merely sensory mental life need not be immediately undercut by the acceptance of a second level of determinable, public, and objective spatiotemporal realities, so too the realities of this second level need not be thought to be immediately undercut by the introduction of a third level of things in themselves. To warrant the undercutting charge here, one would first need to show that there could be no

¹⁵ See Kant's discussion of sensation at A20/B34 and A28/B44.

¹⁶ A19/B33, from the first paragraph of the Transcendental Aesthetic. On the complex relationship between the starting points of Kant's philosophy and common sense, see Ameriks 2003, 1–48 and 2006, 108–33. This interpretation is compatible with saying that it is also worthwhile, as part of the Transcendental Turn, to explain in various ways how starting from a basis of mere sense data and associations would not provide sufficient means to account for basic features of our experience.

relevant and weighty but nonetheless non-demoting notion of things characterized in such an unusual way—and this can hardly be done in advance of a consideration of non-self-defeating ways in which Kant may have wanted technical terms such as ‘in itself’ and ‘ideal’ to be understood. Furthermore, in trying to understand these terms, it is not proper simply to presume, as many interpreters may still do, that the *Critique* employs some kind of ‘short argument’ that puts things in themselves immediately and totally beyond our ken simply because of some *extremely general* feature of our cognition, such as its involving receptivity or representations at all, or concepts that we in part construct and/or create.¹⁷ This would be tantamount to supposing that the very notion of the transcendently real must be that of an item which cannot possibly be represented or conceived at all (because in that way it would be, trivially, related to us rather than simply ‘in itself’), and a fortiori, must be unknowable—a supposition that would immediately leave no room for Kant’s repeated commitment to making room for thinking things in themselves and even having acceptable (nontheoretically justified) beliefs about them.¹⁸

A similar important point to keep in mind here is that, just as for Kant the affirmation of second-level public realities is by no means presented as an *inference* from a base of a first level of merely psychological realities, so too we need not suppose that the mere existence of the items that Kant allows at a third and metaphysical level is to be thought of as a fact originally introduced by means of inference and as playing anything like a typical empirical explanatory role. There are a number of reasons why, given Kant’s ‘anti-Cartesian’ Transcendental Turn against earlier modern philosophy, determinate objective (i.e. ‘second-level’) features should not be understood as inferred from merely subjective mental states: phenomenologically, he does not suppose that we are originally aware of such states as such; epistemologically, he sees that these states are, in any case, in principle too indeterminate by themselves to sustain the relevant inference because, as noted earlier, they are not at a cognitive level and so cannot be part of a genuinely argumentative connection; and, methodologically, he realizes that, whenever we as philosophers do come to assert anything about such states, it is only after having already understood how to make determinate objective empirical claims. Hence, as Kant argues in the Second Analogy and the Refutation of Idealism,

¹⁷ The significance and impact of the ‘short argument’, especially because of Reinhold’s influence, is discussed in Ameriks 2000c and 2005. G. E. Schulze compounded problems here by discussing the *Critique* under the heading of the ‘Kantian-Reinholdian’ philosophy and its supposed basis in the mere concept of representation. Schulze’s book misses Kant’s anti-Cartesian turn and assumes that the Critical philosophy holds to the ‘way of ideas’ and its questionable presumption that ‘representations are the only thing about whose actuality all philosophers agree’, where ‘representation’ is from the start understood as such that ‘the connection between our representations and the things outside us must be *established* above all by a careful and sound answer’ (Schulze 1792, 107). I emphasize ‘established’ to indicate that this is a key term for Schulze, one that he uses as tantamount to satisfying a demand for a justified existential *inference* to something which has nothing to do with representation. He argues that skeptical objections can always be raised against such an inference—as if Kant did not appreciate this point.

¹⁸ See Chignell 2007.

and in contrast to the presumptions of the 'Cartesian skeptic', it is from our grasp of an 'objective sequence' (of determinate physical states: the boat here, then there) that we 'derive', or form and justify, the explicit notion of a 'subjective sequence' (of determinable psychological states: this judgment of appearance, then that judgment), rather than vice versa.

For analogous reasons, Kant's notion of a thing in itself underlying what perceptually appears to us need not be understood as introduced or defined in terms of the notion of something that is totally beyond whatever we are acquainted with and that therefore always has to be inferred; it can simply be the notion of a thing—and, in this case, even something that we are directly acquainted with demonstratively—insofar as it also needs to be understood as having properties that exist in some manner (not positively determinable by us) independently of the appearing properties it has which are such that they can exist only relatively and not all on their own. That is to say, treating something as not existing 'in itself' need not be a matter of regarding it as not existing at all, or merely in a first-level private and subjectivist way, but can be simply a matter of regarding it as lacking the special 'higher' features of a 'thing in itself', the most important of which in Kant's discussions is the feature of not being fundamentally conditioned, that is, determined by a reality that is itself of a fundamentally higher level.¹⁹ This interpretative proposal receives striking support not only from Kant's repeated comments about a thing in itself as that which conditions our sensibility but cannot itself be sensibly conditioned, but also from the other main metaphysical context in which Kant uses the highly unusual phrase 'in itself', namely, his practical philosophy's equally fundamental discussion of the notion of an 'end in itself'. That discussion makes clear, in a way that fortunately has not been as heavily disputed by interpreters (although, even among some readers in that context, there remains a similar 'subjectivist' interpretive fallacy in the unfortunate supposition that happiness has no objective value at all for Kant, whereas all he really holds is that it lacks unconditional value), that something that is not an end 'in itself' can still very definitely be a *real* end. It does not have to be demoted from objectivity altogether but can be treated merely as a sensible and fundamentally conditioned item rather than an unconditional and nonsensible end.

Here it is also very important to keep in mind that, although for Kant 'being spatiotemporally sensible', 'being fundamentally conditioned', and being, in a transcendental sense, 'mere appearance' do all eventually go together, this is not a matter of mere definition but is based on specific arguments that he believes show that in fact there must be something so fundamentally conditioned about our particular spatiotemporal kind of determinable sensibility that this forces us to treat the features disclosed

¹⁹ I link the terms 'features' and 'fundamental conditioning' to help bracket the point that, on all accounts, every thing—except God—might, in its bare *existence*, be dependent on a highest being. That general possibility is surely not Kant's specific reason for introducing a philosophical 'mere appearance'/'in itself' distinction, but the distinction needs to remain compatible with that possibility.

to it as 'mere appearance' in the double transcendental sense of being something that appears objectively and yet also points to the higher third-level reality of an underlying thing in itself. Such an appearance is not 'mere' in relation to its first-level and merely sensory correlate, for it is still an objective *Erscheinung*, a thing that appears, and for that reason is higher than *blosser Schein*, which is mere illusion and a strictly subjective state.²⁰ The key point here is that the fact that something is at a higher level than something else does not mean that the latter item must lack objective reality and be illusory or non-existent; it merely means that the latter item is ontologically less basic, and this is compatible with such a second-level feature still playing a genuinely objective and not only central phenomenological role in our experience.

It can still seem very controversial, of course, to claim, as Kant's transcendental idealism does, that spatiotemporal features have to be grounded, in some way that is neither analytically nor scientifically determinable, on more fundamental nonspatiotemporal features, but the main initial worry about such a claim should not be that it would supposedly withdraw all objective reality from our experience but rather that it can involve adding more layers of reality than are indisputably needed. But although this kind of expansive ontological position can admittedly be worrisome in its own way, appreciating its true meaning can also bring with it some interpretive advantages. In particular, thinking in terms of this position can help in understanding how Kant's philosophy has several striking similarities with other related objectivist philosophies that also call themselves 'idealist' in a way that is often deeply misunderstood. Hegel, for example, repeatedly makes clear that when he uses the terms 'idealism' and 'appearance' and calls finite spatiotemporal things 'mere appearance', what he basically means is just that each one of them is fundamentally conditioned by a metaphysical reality that itself is not similarly conditioned, in this case, the underlying total rational structure that he calls the 'absolute'.²¹ The fundamental similarity with Kant here lies in the fact that—contrary to many once common but very skewed readings of his system—Hegel still considers the spatiotemporal and fundamentally conditioned things that appear to us to be objective realities. As essential, rationally structured, and universally appearing parts of the absolute, these entities definitely cannot be demoted to sheer nonbeing, illusion, or psychology. Even though it also remains true that Kant, unlike Hegel, rejects ontological holism and the epistemological view that we can theoretically determine that which fundamentally conditions appearances, these are but incidental points of difference in contrast to the more basic metaphysical and terminological issue of treating physical items as real in a genuinely objective sense while nonetheless using the term 'ideal' for them.

²⁰ Illusions have the complex quality of being like sensations, in being merely subjective states, and yet also unlike them, because illusions inherit the syntactical complexity that characterizes our objective judgmental states, which are correlates of appearances, i.e. ways that things actually appear. See Prauss 1971.

²¹ Hegel 1830, §§45, 50 (*Encyclopedia Logic*). See the discussion in Ameriks 2000a, 276 n. 18, and 2006, 157–8. Hegel himself seems not to have appreciated how close he was to Kant because, under Jacobi, Reinhold, and Schulze's influence, he tended to read Kant's metaphysics in subjectivist terms.

This terminological peculiarity can be made less perplexing if one keeps in mind that in the German tradition, which was heavily influenced by Platonism, the term 'ideal' does not have to carry the primarily negative and merely subjective meaning that it tends to have in English, as if it must be equivalent to 'not actual', but, on the contrary, is often used to stand for that which is most real or is at least directly conditioned by the most real.²² In addition, positions even closer to Kant's, although with somewhat different terminology, can be found in the philosophies of the early German romantics and the late Schelling. They explicitly follow Kant's general metaphysical notion of a distinction between appearances and a thing in itself to the extent that they also insist on the key point that all the spatiotemporal features that we can know are objectively real and yet are fundamentally conditioned by something else—something unconditioned that we *cannot* expect to be able to determine theoretically in the way that Hegel supposes.²³

Spinoza, Jacobi, and even some contemporary naturalists who take holism seriously can all be said to be committed to a somewhat similar position, insofar as they also grant that the particular spatiotemporal features that can appear to us are real and yet are fundamentally conditioned by something else.²⁴ Although Kant would definitely have regarded them as even further from his own view, these still popular options—Spinoza's conception of absolute substance, Jacobi's intuition of God, and recent cosmological theories about ultimate underlying physical grounds of the universe that in some (possibly in principle uncognizable) way 'precede' even the structures of space-time—obviously differ significantly from each other, but, like Kant's philosophy, they all can be taken to imply a coherent, expansive, and multilevel realist metaphysics.

Whatever basic similarities and incidental differences arise here concerning the *conclusions* about the specific nature of the nonempirically conditioned reality that all these metaphysical views introduce, there remain even greater differences concerning the kinds of premises and *arguments* that are used to try to reach these conclusions. Nothing that has been said here so far can serve to begin to *justify* the highly distinctive and still mysterious specific arguments for idealism that Kant offers in the Aesthetic and the Antinomies. This serious difficulty may even be what silently underlies the desire of most contemporary interpreters to keep the Transcendental Turn free from entanglements with metaphysics—but if that is the case, it becomes all the more worthwhile to uncover the surprising extent to which such interpreters have repeatedly focused instead on other, much more extreme and controversial objections. And even though this also implies that so far we have been only beginning to outline the project of satisfactorily reconciling Kant's own notion of transcendental idealism (or at least of genuine analogues to it) with his much more widely embraced notion of a Transcendental Turn, it is important, nevertheless, to make a beginning and finally gain some sense of what such a project would truly need to involve.

²² See Ameriks 2000b and Beiser 2003.

²³ See Ameriks 2000b and 2006, 149–60.

²⁴ These themes are explored in very helpful detail in recent work by Paul Franks, especially Franks 2005.

Given all the metaphysical cousins of Kant's project that have just been uncovered, the mere idea of approaching his idealism in terms of a multilevel ontological approach that is consistent and primarily additive rather than contradictory and/or absurdly subtractive should at least no longer seem so unusual or inherently implausible. This can remain true even though at one key point Kant's position admittedly sounds subtractive insofar as it involves explicitly rejecting the transcendental reality of space and time (and of all cognition that depends entirely on the features of space and time). Even this rejection can be understood in a relatively harmless way, however, as by itself merely a denial of *inflationary* positions that ascribe a kind of unconditioned status to items that cannot be unconditioned in a relevant metaphysical sense, and so, as Kant repeatedly stresses, it is not a 'material' rejection of the existence of empirical objective items as such, let alone of independent reality altogether.²⁵ This makes it all the more vital, however, to be clear about properly identifying the relevant 'inflationary' claim about empirical objective items that is being rejected when Kant denies, on this ground, that such items can ever be characterized as things in themselves.

This point is relevant to a recent argument by Henry Allison, who proposes that to affirm the 'transcendental reality' of some feature is to say that *all* possible things *must* have that feature.²⁶ He then adds that, because it would be dogmatic to hold to the metaphysical assertion that all things must be spatiotemporal (for what do we know about what other realms and 'possible worlds' there really are or might have been?), it immediately and harmlessly follows that we should not say that space and time possess transcendental reality. Moreover, given this specific definition of transcendental reality, it is easy to go further and deny *any* positive assertion about things in themselves as such, for there is no reason to suppose that we can determine any specific kinds of concrete things that must be found in all 'possible worlds'.

This reading is very much in the tradition of the extreme antimetaphysical interpretation, but there are problems with taking it to be a successful new way of establishing an anodyne meaning for Kant's transcendental idealism and his correlative denial of the transcendental reality of empirical properties. It is hard to see how it can work as a general reading of Kant's doctrine because it is essential, for reasons already clearly indicated in the first *Critique*, that the properties of being a soul and being absolutely free (whether or not we can theoretically justify their actual existence), for example, could be transcendently real without having to be thought of as applying to all transcendental realities, that is, all things in themselves. The reading has an understandable source, to be sure, in the thought that *one* meaning of 'transcendental' is 'holding in all contexts', that is, categorially. But this does not show that Kant's actual use of the term 'transcendently real' in contrast to 'transcendently ideal' is such that by definition it could apply only to features that are themselves like categories in necessarily characterizing all things that might exist; on the contrary, such a definition would directly

²⁵ See esp. Kant's 'Refutation of Idealism', B274–5, and his *Prolegomena*, discussed in section 6.

²⁶ Allison 2006. This argument goes beyond the earlier interpretation in Allison 1983/2004.

conflict with repeated distinctions that Kant makes in dividing the different sections of the Transcendental Dialectic. Moreover, although in denying the transcendental reality of something Kant could *sometimes* be making a nonsubstantive claim that does not imply the transcendental reality of anything else at all, this does not mean that his transcendental idealism cannot involve a commitment to a positive, even though necessarily indeterminate, metaphysical assertion of something that is transcendently real and nonsensory. On the contrary, the natural way to read Kant's repeated remarks about space and time as *merely* ideal, and his repeated remarks that anything *transcendentally* characterized as an appearance requires *some thing* that appears,²⁷ is to take them to show that he believes that whenever we are forced to treat certain features as fundamentally conditioned, we then must also allow that there remains something that does the conditioning, namely a thing that is 'in itself' and is not conditioned in the same sense.

It might seem dogmatic to insist in this way on something unconditioned, and yet for Kant's system such an insistence appears unavoidable.²⁸ The alternative would be to say that the appearances that confront us are themselves without a condition—and to say this would then still be to posit something unconditioned, namely, the spatiotemporal appearances themselves. This option, however, would involve the horrendous situation of calling precisely those items unconditioned that—given Kant's key First Antinomy arguments—have the demonstrable feature of being unable to be unconditioned, no matter what other level of reality they may nonetheless have. Hence, given that on Kant's arguments for idealism there is something fundamentally conditioned about the features of appearance, it does follow that he requires something in itself that is not appearance and that grounds those features. This is, however, only a very general and indeterminate bare existence claim, one which does not positively determine the intrinsic nature of any such thing, and hence it need not be thought to introduce a deeper structure that would demote the objective reality of the structure of our determinable spatiotemporal world—any more than the even richer thought and assertion of unconditioned causality in a free agent in itself would demote the reality of the empirical effects of that agent.

The notion of things in themselves as conditions for fundamentally conditioned items has also been stressed in a very challenging new kind of metaphysical interpretation of Kant's idealism by Rae Langton.²⁹ Her reading is a version of

²⁷ See e.g. A251, Bxxvi–xxvii, B306, A494/B522.

²⁸ On reason's 'rightful' demand for the unconditioned in things in themselves, see Bxx. The sense of 'unconditioned' that matters here can be for something that is merely unconditioned in some nonempirical way, and that item need not be unconditioned in all respects. Instances of noumenal freedom or noumenal sensory affection can thus be called unconditioned, because of their not themselves being grounded in a particular way, and yet they could be dependent in other ways. What is most important here is to be clear about the *transcendental* notion of a 'mere appearance', which is the notion of something that requires a nonempirical ground *and* that itself sensibly appears. Hence, even if the nonsensible ground of such an appearance turns out to be in some way itself conditioned, this ground would not thereby become a 'mere appearance'—because it is not something that sensibly appears.

²⁹ Langton 1998.

a Moderate Interpretation insofar as it also expresses a kind of overall realist and ‘non-demoting’ approach, in this case by proposing that Kant’s things in themselves are to be understood entirely in terms of a thing’s underlying nonrelational and intrinsic properties. The existence of such properties can explain how a thing in itself can be characterized as having a fundamentally more basic level of reality without threatening the objective reality of the dependent items that it conditions. Her proposal is, nevertheless, unlike the particular kind of Moderate Interpretation that I have been presenting because of the way that it neglects Kant’s specific Critical arguments for transcendental idealism. These arguments depend essentially on specific problems with taking space and time to characterize things in themselves, rather than, as on her proposal, a version of the ‘short argument’ that would bypass this issue and simply take our general receptivity and inability to reduce relations to be sufficient to entail ideality.³⁰ Because her reading must rely largely on points from texts that antedate even Kant’s 1770 *Dissertation*, it misses the distinctive premises behind the *Critique*’s idealism. It is also not clear how the reading can do justice to the key conclusions of Kant’s full system and allow it to be even possible for our free causings or other crucial relational truths—such as God’s power in helping to enable the highest good—to belong, as Kant evidently intends, to the level of things in themselves.

Despite the virtues of its realism, Langton’s reading also minimizes the *specific sensory* nature of our receptivity, which plays a key role in the versions of the Moderate Interpretation that I believe are most promising, namely, recent versions that build on an analogy between Kant’s idealism and contemporary ‘realist’ readings of the qualities of our perceptual experience.³¹ On these kinds of readings, color, for example, is understood neither as a first-level item, a merely subjective and psychological state, nor as a third-level feature of objects, for such a feature would have to be able to be adequately described entirely apart from any reference to our general kind of sensibility.³² On the new realist readings, colors are understood as simultaneously subjective and objective, for they standardly involve both an actual way that a thing *phenomenally appears* to the minds of beings like us, and also a kind of disposition *belonging to the thing* as an empirical object, that is, with a specific kind of surface that in a specific kind of circumstances always would appear in a specific kind of way.³³ These dispositions in turn call for some kind of underlying categorical base, and it is at this point that the Kantian notion of the features of things in themselves can be introduced as *analogous* (but only analogous) to the ultimate—but also very hard ever to pin down—categorical (and

³⁰ Ameriks 2003, 135–57 (‘Kant and Short Arguments to Humility’).

³¹ See Allais 2007 and Rosefeldt 2007.

³² Note that these readings are not meant to be based on Kant’s own specific and shifting characterizations of secondary qualities as such but are developed as valuable for purposes of analogical thinking, which always involves appreciating points of dissimilarity as well as of similarity.

³³ There are significant variations in discussions of such theories. See e.g. McDowell 1998, 112–50, McGinn 1999, 324–5, and Stroud 2000.

non-appearance-involving) properties that even contemporary naturalist metaphysicians often insist on.

It is not just any set of items that are involved with dispositions, however, but realistically construed perceptual qualities in particular that provide the best examples of what the Moderate Interpretation means by a level that is 'in between' the entirely subjective and the entirely objective realms. In Kant's terms, this realm of empirical reality or 'transcendental ideality' lies between domain of the merely subjectively real or 'empirically ideal' and the domain of the 'transcendentally real' or 'in itself'. Here the analogy with perceptual qualities is a helpful reminder that there are two conditions that need to be satisfied when calling something an appearance, that is, a 'mere appearance' in Kant's *transcendental* sense: first, that, as 'appearance', it actually involve something appearing in a sensory and objective way,³⁴ and, second, that, as 'mere', it point to, or 'signify', something fundamentally (and not merely empirically) more basic.

5. Kant's Alleged Incoherence: A Bad Causality Interpretation and Subjectivism

It is impossible in a limited space to defend the Moderate Interpretation against all the extreme alternatives to it, but a fair indication of how such a defense would go can be gained from considering possible responses to the aggressive attack that Allen Wood carries out against what he calls 'the Causality Interpretation'. Wood presents this interpretation as one of the two ways that Kant's idealism can be understood. The first option, the 'Identity Interpretation', ontologically 'identifies' appearances with things in themselves and denies they are causally connected, whereas the second option, the Causality Interpretation, distinguishes appearances and things in themselves and relies on what Wood considers to be Kant's highly unfortunate discussion of causal relations in passages such as A30/B45 and A43/B60, which speak straightforwardly about things in themselves as causes distinct from 'our' representations.³⁵

Although Wood expresses his objections here fundamentally in terms of the issue of causality, the most striking feature of his account of the Causality Interpretation is that he immediately characterizes the representations that it involves—in contrast to the account of the Identity Interpretation—not at all in terms of objective appearances but instead in terms of 'merely subjective entities or states of mind' in us.³⁶ In other words, he implies that a causality account must treat Kantian representations merely as representings or sensations, rather than ever as objective representeds, although this

³⁴ I take it that Kant (A226/B273) does not require that an empirically objective feature has to be something that human beings, given the 'grossness of their senses', *can* directly 'see', but only that it is the kind of item that they *could* see, without changing into a fundamentally different kind of being, in an external situation that allows for such seeing.

³⁵ Wood 2005, 64.

³⁶ Wood 2005, 65.

immediately erases the crucial distinction that Kant repeatedly makes, and that the Moderate Interpretation stresses, between these two levels.

It is precisely at this point that Wood makes his key claim that the Identity Interpretation and supposedly it alone, does not 'demote' phenomena from reality to ideality—as if the ideality that any causality account might involve would have to be tantamount to a reversion to psychological idealism and an espousal of Jacobi's uncharitable reading of Kant.³⁷ Here it is not merely the thing in itself but the very idea of even an empirical object that has become entirely mysterious, simply because from the very start a distinction between dependent public appearances and mere private sensations has been disallowed. Wood also contends that the only alternative to his subjectivistically construed Causality Interpretation would be an Identity Interpretation that simply insists, as he says, that 'the bodies we cognize have an existence in themselves'.³⁸ But this insistence, with its direct application to bodies of the phrase 'in themselves', does not clearly match any passages that Wood cites, and it clearly conflicts with texts that he does cite. By neglecting the Moderate Interpretation's alternative of treating appearances as an 'in between', simultaneously non-in-itself and non-subjectivistic form of dependent existence, Wood excludes the possibility of a consistent and broadly realistic reading of both appearances and (grounding) things in themselves, one exemplifying a coherent form of what could still be called a kind of causality account of idealism.³⁹

³⁷ Wood 2005, 65. At Wood 2005, 64, Wood cites a passage from A387 that may have misled him because it speaks of bodies and motions as both 'representations in us' and 'equally' grounded in things in themselves, as if there are no empirically external objects and as if there is only noumenal but not empirical causation (because there would be no bodies around to cause motions). Here it is important to recall Kant's clarification at the very beginning of the section (A367) that he will be making use of an 'intellectual' or transcendental rather than merely empirical distinction between what is 'inside us' and what is 'outside'. This means that something's being 'inside' us can signify simply that it does not belong outside of us in things 'in themselves', and yet it still could be outside us in an empirical sense. In the A387 passage, Kant can be taken to be simply pointing out, along the lines of the Transcendental Turn, that any specific causal relation we can determinately know concerns the transcendental 'in us' realm and so cannot make claims about what is 'entirely heterogeneous' to us in that sense, i.e. about properties of a thing in itself, which as such are (given the *Critique's* earlier arguments) unlike anything within our experience. Kant's view does not exclude, I believe, the *indeterminate* philosophical claim that there is something nonempirical that affects us, because that claim does not positively determine what the affecting thing is like and how it operates.

³⁸ Wood 2005, 66.

³⁹ Wood also insists that a Causality Interpretation has to imply that cause and effect are distinct entities (Wood 2005, 65), and that this supposedly goes against a well-known passage at Bxxvi–xxvii, which speaks of 'objects of appearance and the very same things as things in themselves'. But the overriding context of that passage, and the only concrete example given there, clearly concerns human freedom (Bxxviii), which is a very special case for Kant. There are unique reasons in Kant's view that call for us to have to be able at least to think in terms of identity here: Kant insists that our inescapable conception of moral personality demands that our notions of a thing in itself and an appearance must be such that this conception can be both theoretically thought of and, on pure practical grounds, positively applied in regard to the very same individual. The specific passage can hardly be taken as proof, however, of a general claim that Kant must mean the very strong claim that the same individuals must *always* be involved as both things in themselves and appearances, for (aside from opening the door to improper detailed knowledge of noumenal individuation) that would make little sense e.g. of his discussions of God, who Kant clearly allows to be considered as a thing in itself but not a phenomenon.

6. Back to Kant: The Text Itself

One way to try to begin to settle these complex interpretive issues is to examine Kant's texts more directly, and to look in particular at Wood's main exegetical argument that Kant falls into 'nonsense and self-contradiction' and 'Orwellian doubletalk' when characterizing and defending his own idealism.⁴⁰ Wood emphasizes the following passage from Kant's *Prolegomena*, First Part, Note 2:

[Against idealism] I say that things as objects of our senses existing [situated] outside ourselves are given, but we know nothing of what they may be in themselves, cognizing only their appearances, that is, the representations that they cause in us by affecting our senses. Consequently, I grant by all means that there are bodies outside us, that is, things[,] which, though quite unknown to us as to what they are in themselves, we still cognize by the representations which their influence on our sensibility procures us, and which we call bodies, a term signifying the appearance of the thing which is unknown to us but not the less actual.⁴¹

Wood begins his analysis by claiming that Kant in 'the first sentence here says that objects of our senses are given to our cognition, but then denies that we cognize these objects'.⁴² But, surely, a more natural, and moderate, reading of the first sentence is that what Kant explicitly says his doctrine denies here is just that we cognize what these objects 'may be in themselves'. Similarly, it is not necessary to insist that when Kant says that what we do cognize is 'their appearances', although these are 'only' appearances, this entails any denial (*pace* Wood) that we know the 'objects of our senses' that the sentence begins with. In other words, in this transcendental context, the notion of what is 'only appearance' is most naturally to be understood as marking a contrast to the immediately preceding term 'in themselves'.

Wood goes on to claim that the second sentence here shows Kant 'infers' that 'there are bodies outside us', and yet Kant supposedly 'proceeds to say', absurdly,

⁴⁰ Wood 2005, 67.

⁴¹ Wood 2005, 67, citing Kant 1783 (4: 287). I have supplemented the translation in Wood with the additions in brackets; only in the case of the comma do I believe that the correction of a misleading impression may be involved. The paragraph concludes: 'Can this be idealism? It is the very opposite of it.' At the outset of the paragraph Kant clearly defines the version of 'idealism' that he takes to be directly rejected on his philosophy: 'Idealism consists in the assertion that there are none other than thinking beings.' My hypothesis is that Wood and other readers might nonetheless be misled by this paragraph because of the fact that it speaks of our knowing 'by representations', and of the appearances we know as 'i.e. representations' and something 'caused in us'. By itself this might sound like the subjectivism explicitly rejected from the start of the paragraph, which would be odd and incoherent. But the key terms in this section are ambiguous (see my n. 37, concerning transcendental and empirical meanings for 'in'), and even an ordinary empirical realist could express himself similarly. Although elsewhere (e.g. B7 on.) Kant is clearer about how the term 'appearance' can, and generally should, be used in an objective as opposed to merely subjective sense, i.e. as designating the way that an object truly appears, rather than merely an event of appearing within us, there is no need to try to get around the fact that the term also always involves something subjective. The main question is just whether the representations that are mentioned in this passage *can* also be understood as involving an appearance in the objective sense of a genuine cognition of an outer spatial thing—a question that this paragraph repeatedly answers affirmatively.

⁴² Wood 2005, 67.

‘that it is not these bodies that we call “bodies”, but rather *bodies* are a wholly different set of entities’.⁴³ It is not at all clear, however, that Kant’s text needs to be read in this way as absurd doubletalk. Kant can be taken to be simply making, in only slightly different terms, the same basic point over again that he made right from the start: the sentence reiterates literally that ‘there are bodies outside us’, which we ‘still cognize’, although they remain ‘quite unknown to us as to what they are in themselves’.⁴⁴ The second sentence does have the advantage of making clearer that our knowledge occurs ‘by representations’, that is to say, ‘through’ them as acts of mind, and, one can even add (to make sense of Kant’s earlier phrase ‘appearances, that is, the representations’), it is also true, in a *transcendental* sense, that our empirical knowledge consists in, and is only of, ‘representations’, in the sense that what we know, even as objective, is always related to representations of a specific relative and dependent manner of being and thus cannot be said to apply to things as they are simply ‘in themselves’.

Here Kant does not need to be said (*pace* Wood) to be ‘inferring’ bodies from anything else, for the second sentence can be read as simply drawing an elementary analytic consequence of the prior sentence, in which Kant had already declared, against his subjectivist interpreters, that he (like the rest of us) does hold that there are ‘objects of our senses existing outside ourselves’.⁴⁵ Thus, there are bodies, for that is what bodies are, namely sensible objects (and not mere acts of sensing within us)—although it remains true, given Kant’s philosophy, that these bodies as such are not things in themselves but can only, as he reminds us in conclusion here, ‘signify the appearance’ of ‘unknown to us’ things in themselves. Once again, we need not suppose (*pace* Wood) that at the end Kant is saying that *bodies* are in any sense a ‘wholly different set of entities’ than what we, in our sensible cognition, call bodies. We can understand Kant as simply saying again that, *in addition* to the fact that there are outer sensible entities that we do know, and which are understandably called bodies, there is the fact of things unknown ‘as to what they are *in themselves*’ (*italics added*).

This is all, by rough analogy, not much more mysterious than children at a zoo saying that, in addition to an animal body surface that they see, there is an unknown set

⁴³ Wood 2005, 67.

⁴⁴ Kant’s text adds, right at the beginning of the next paragraph after Wood’s quote, that, despite our noumenal ignorance, we can say that what is outer is not only ‘given’ through sensibility but also ‘falls under many of the predicates of outer things’ (4: 289). It also adds that, although these predicates are valid, they lack existence ‘on their own’, that is as characterizing things ‘in themselves’. Presumably this means we cannot understand them without a possible relation to appearing in a certain way to human sensibility. In questioning just the transcendental reality of these predicates, Kant should not be assumed to be denying the empirical reality of bodies (let alone their real external but nonempirical ground)—any more than we should assume that persons must be denying the reality of colors just because they do not accept that something can be a color simply ‘on its own’, i.e. as a color, and yet such that it *could not* appear as such to anything like human sensibility.

⁴⁵ Kant 1783 (4: 289).

of inner features, a set in part responsible for enabling that body to present itself. In a loose sense, the 'inside' of the animal is identical with, that is, belongs to, the same entity as the animal's surface, and yet, in another sense, the animal's surface (as well as our perceiving of it) is in various ways both distinct from and the consequence of the animal's inner features. The only methodological difference here is that Kant's 'inner' features are introduced for philosophical rather than elementary scientific reasons, and they and their specific inner operations are unknowable in principle, rather than accidentally, because of specific arguments given elsewhere about the limitations of spatiotemporality. As noted earlier, Kant's specific arguments may certainly be questioned, but Wood's attack here is aimed at the mere meaning and consistency of Kant's assertions, and it is infected by the extremely uncharitable assumption that a denial of knowledge of what a body is 'in itself' has to amount to a denial of our knowledge of the body's objective predicates and reality altogether.⁴⁶

One should now be able to see how, *given* their own assumptions, interpreters such as Wood can come to characterize as 'patent nonsense from the critical point of view' any interpretation that holds to 'the letter' of those texts in which Kant continues to affirm that things in themselves affect our sensibility.⁴⁷ And now one should also be able to see how the Moderate Interpretation offers a relevant alternative to these assumptions, one that can help to preserve the empirical realism of the Transcendental Turn and save readers from having to say that the mature Kant falls into 'patent nonsense' even from his own point of view.

References

- Allais, Lucy (2007), 'Kant's Idealism and the Secondary Quality Analogy', *Journal of the History of Philosophy*, 45, 459–84.
- Allison, Henry E. (1983/2004), *Kant's Transcendental Idealism: An Interpretation and Defense*, revised and enlarged edn (New Haven: Yale University Press).
- Allison, Henry E. (2006), 'Transcendental Realism, Empirical Realism, and Transcendental Idealism', *Kantian Review*, 11, 1–28.
- Ameriks, Karl (1982/2000a), *Kant's Theory of Mind* (Oxford: Clarendon Press).
- Ameriks, Karl (2000b), 'Introduction: Interpreting German Idealism', in Karl Ameriks (ed.), *The Cambridge Companion to German Idealism* (Cambridge: Cambridge University Press), 1–17.
- Ameriks, Karl (2000c), *Kant and the Fate of Autonomy: Problems in Appropriating the Critical Philosophy* (Cambridge: Cambridge University Press).
- Ameriks, Karl (2003), *Interpreting Kant's Critiques* (Oxford: Clarendon Press).

⁴⁶ Perhaps the ultimate—and undeveloped—worry here is not a subjectivist, or causal, or methodological one, but just that the very notion of nonspatiotemporal properties is incoherent. This claim might still be true, but it is an oddly dogmatic claim (somewhat like early objections to the very notion of action at a distance) to insist on, especially in an era in which countless rigorous theories of psychological, modal, moral, and even cosmological properties are being developed without any such dogmatic presupposition.

⁴⁷ Wood 2005, 70.

- Ameriks, Karl (2005), 'Introduction', in Karl Leonhard Reinhold, *Letters on the Kantian Philosophy*, ed. Karl Ameriks (Cambridge: Cambridge University Press), ix–xxxv.
- Ameriks, Karl (2006), *Kant and the Historical Turn: Philosophy as Critical Interpretation* (Oxford: Clarendon Press).
- Beiser, Frederick (2003), 'Two Concepts of Reason in German Idealism', *Internationales Jahrbuch des Deutschen Idealismus/International Yearbook of German Idealism*, 1, 15–27.
- Bird, Graham (2006), *The Revolutionary Kant: A Commentary on the Critique of Pure Reason* (Chicago and La Salle, IL: Open Court).
- Chignell, Andrew (2007), 'Belief in Kant', *Philosophical Review*, 166, 323–60.
- Franks, Paul (2005), *All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism* (Cambridge, MA: Harvard University Press).
- Gardner, Sebastian (2005), 'Review of Henry E. Allison, *Kant's Transcendental Idealism: An Interpretation and Defense*, Revised and Enlarged Edition', *Notre Dame Philosophical Reviews*, 4 Sept.
- Hegel, G. W. F. (1830), *Enzyklopädie der philosophischen Wissenschaften im Grundrisse, Erster Teil*, 3rd edn (= *Encyclopedia Logic*) (Heidelberg: C. F. Winter).
- Jacobi, Friedrich Heinrich (1787), *David Hume on Faith, or Idealism and Realism: A Dialogue*, in *Friedrich Heinrich Jacobi: The Main Philosophical Writings and the Novel 'Allwill'*, ed. and tr. George di Giovanni (Kingston and Montreal: McGill University Press, 1994).
- Kant, Immanuel (1781/1787), *Kritik der reinen Vernunft* (= *Critique of Pure Reason*) (Riga: Hartknoch).
- Kant, Immanuel (1783), *Prolegomena zu einer jeden künftigen Metaphysik die als Wissenschaft wird auftreten können* (= *Prolegomena*) (Riga: Hartknoch). This and other works by Kant, with the exception of the *Critique of Pure Reason*, are cited by volume and page in *Kant's gesammelte Schriften* (Berlin: de Gruyter, 1900–).
- Kant, Immanuel (1790a), *Kritik der Urteilskraft* (= *Critique of the Power of Judgment*) (Berlin and Libau: Lagarde and Friederich).
- Kant, Immanuel (1790b), *Über eine Entdeckung, nach der alle neue Kritik der reinen Vernunft durch eine ältere entbehrlich gemacht werden soll* (= *Contra Eberhard*) (Königsberg: Friedrich Nicolovius).
- Langton, Rae (1998), *Kantian Humility* (Oxford: Oxford University Press).
- McDowell, John (1994), *Mind and World* (Cambridge, MA: Harvard University Press).
- McDowell, John (1998), *Mind, Value, and Reality* (Cambridge, MA: Harvard University Press).
- McGinn, Colin (1999), *Knowledge and Reality: Selected Essays* (Oxford: Clarendon Press).
- Prauss, Gerold (1971), *Erscheinung bei Kant* (Berlin: de Gruyter).
- Rosefeldt, Tobias (2007), 'Dinge an sich und sekundäre Qualitäten', in Jürgen Stolzenberg (ed.), *Kant in der Gegenwart* (Berlin: de Gruyter), 167–209.
- Sacks, Mark (2000), *Objectivity and Insight* (Oxford: Clarendon Press).
- Schulze, G. E. (1792), *Aenesidemus or Concerning the Foundations of the Philosophy of The Elements Issued by Prof. Reinhold in Jena Together with a Defense of Skepticism Against the Pretensions of the Critique of Reason*, tr. in part in George di Giovanni and H. S. Harris (eds.), *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism*, rev. edn (Indianapolis: Hackett, 2000), 104–35.
- Strawson, P. F. (1966), *The Bounds of Sense* (London: Methuen).

- Stroud, Barry (1999), 'The Goal of Transcendental Arguments,' in Robert Stern (ed.), *Transcendental Arguments: Problems and Prospects* (Oxford: Clarendon Press), 155–72.
- Stroud, Barry (2000), *The Quest for Reality: Subjectivism and the Metaphysics of Color* (New York and Oxford: Oxford University Press).
- Wood, Allen W. (2005), *Kant* (Malden, MA: Blackwell).

3

Kant, Naturalism, and the Reach of Practical Reason

Paul Abela

Cordelia, Cordelia, stay a little.
(*King Lear*)

When moral theorists are looking for a secure backstop against the creep of contemporary naturalist accounts of practical reason, Kant's approach is never far from the stage. After all, it is the Kantian analysis that locates the good life firmly within the boundary of reason. The categorical imperative directs attention away from naturalist categories of embodied interest, towards rationalist considerations of disinterested obligation. In place of the contingent push and pull of desire satisfaction models, the Kantian analysis enforces requirements of universality and necessity. Transcendental considerations, not empirical conditions, frame the structure of the good life.

This picture of Kant's transcendental turn is as false as it is familiar. In reality, Kant's mature transcendental theory of practical reason offers a nuanced, and tangled, relation to modern naturalism. While Kant's treatment of obligation does point away from naturalistic accounts of moral motivation, Kant's abiding framework for action—the task of realizing the highest good—directs practical attention towards the empirical conditions that support the struggle to make ourselves 'worthy of happiness'.

By locating the following discussion in the context of Kant's larger description of practical reason, it is hoped that some otherwise submerged themes can be brought to the surface. For example, contemporary research from the emerging science of happiness, with its careful attention to the causal conditions of flourishing, is found to be more friend than foe. Moving in the other direction, Kant's transcendental analysis of the contingent alignment of the demands of the good life with the disinterested

deliverances of nature reveals a sensitivity to the costs of moral struggle that finds no counterpart in naturalism.

This chapter begins with a sketch of the received position, reviewing the standard anti-naturalist reading of Kant and introducing a working description of naturalism. Section 2 offers reasons for favouring the highest-good model as the preferred architecture for framing Kant's transcendental account of practical reason and highlighting the points of convergence between his account of happiness and present-day findings. Sections 3 and 4 move in the opposing direction, finding elements in Kant's analysis that are wanting in the naturalist characterization of the conditions of the good life. Section 3 highlights how Kant's conception of obligation serves as a cautionary tale against reductivist naturalist accounts. Section 4 addresses a rarely discussed aspect of Kantian moral experience: the significance of the tragic space that lies between our 'wish and will', and the deliverances of disenchanted nature. While seemingly a distant concern for both Kantians and naturalists, on the proffered reading, Kant's transcendental analysis is shown to offer a surprisingly poignant appreciation of the empirical reality of tragedy. Naturalism, by contrast, is exposed as blind (deaf and silent) to the significance of this reality.

1. The Contemporary Playing Field

The Kantian account of practical reason is read, with few exceptions, as anti-naturalist. Kant's sharp distinction between questions of fact and matters of obligation, of 'is' and 'ought', leaves little room for common naturalist attempts to derive moral obligation from self-standing natural drives or patterns of law-determined social relations. Not surprisingly, the Kantian account of moral agency is the standard whipping boy for naturalists of all stripes. When Daniel Dennett derides the Kantian moral approach as a rationalist fantasy that conceals the better path, he speaks for many.¹

'Naturalism' has a great arc of employment in contemporary discourse. It is often a stand-in for methodological principles perhaps better framed under evidentialist forms of empiricism. It is sometimes used as short form for the abiding evolutionary framework, focusing the understanding of human beings, and their place in nature, through the lens of blind causal forces operating under natural selective pressure. Philosophical naturalism espouses, to varying degrees, a commitment to physicalism, viewing physicalism as the appropriate framework for satisfying legitimate forms of explanation.²

¹ Daniel Dennett, *Freedom Evolves* (New York: Penguin Books, 2003), 213. As Dennett is an articulate and outspoken advocate for naturalism, I press him into service throughout. One could add many other contemporary naturalists who share the view that a scientific understanding of human behaviour will (and should) displace the kind of transcendental analysis associated with Kant. At the loud end of evolutionary psychology see Richard Dawkins's work. For a more nuanced view, see Steven Pinker, *The Blank Slate* (New York: Penguin, 2002).

² Most naturalists are, minimally, methodological naturalists. They look to the canons of evidence—appeals to objects and their properties—as the basis for explanation and prediction. Physicalism, strictly construed, is a metaphysical doctrine that claims that every real thing is a physical thing.

For the purposes of highlighting the points of contact with Kant, 'naturalism' is best understood as a doctrine that employs an empiricist methodological principle within a physicalist metaphysics, registering an insistence that explanations are valid only in so far as they appeal to the properties (known or hypothesized) of empirical objects. In this way, objects, and the causal laws that quantify and calibrate their properties and interactions, provide the template for understanding all phenomena, including human behaviour.

This rough description leaves open intramural debates concerning realism, intuitionism, constructivism, etc. What is important is that naturalists insist that the methodological canons of evidence, and the ontological priority of physical properties, exhaustively orient what can count as legitimate forms of explanation. This approach need not model naturalism as a reductivist programme. One could be a naturalist in the required sense and still affirm, for example, the supervenience of the mental. Or one might look to biological categories (e.g. natural selection) as non-reducible emergent properties. What naturalism does exclude is any appeal to non-physical causal elements for purposes of explanation or prediction; effectively substituting causes for reasons, in addition to banishing all appeals to supernatural causation.

Against this backdrop, it is an incontestable truth that Kant's analysis locates moral value outside the space of causal engagement. The various formulae of the categorical imperative express, in their different ways, one truth. Self-consistent reason use among moral agents imposes constraints on action. Autonomy is understood as the free expression of rational judgement subject to the constraint imposed by duty as revealed by the categorical imperative. The capacity to impose a law upon oneself stands at the centre of this approach to agency and moral value.

So construed, this familiar deontological orientation invites the standard divorce of moral value from an evaluation of consequences *in toto*—never mind particular naturalist categories of consequences in terms of appeals to selfish genes, defecting/non-defecting utility maximization, group evolutionary strategies, etc. In short, the Kantian view appeals to the structure of moral judgement, to reasons over causes (physiological, psychological, anthropological), as the non-reductive ground for locating and assessing the moral status of action. This is the familiar picture of the Kantian transcendental turn in practical reason, evident to all with even a passing acquaintance with the *Groundwork*.

These themes are not surprisingly anathema to a naturalist account of human flourishing. The appeal to reasons over causes, the non-empirical standing of the categorical imperative, the appeal to moral motivation in terms of an intellectual affect primed by the concept of duty (respect), even the occasional invoking of noumenal causation, all point to an utter rejection of naturalism.

This is where the story usually ends, not just for naturalist, but for many moral theorists (Kantians included) as well. As much as naturalists need and find a strawman in Kant, many Kantians view him as a useful bulwark against contemporary attempts to denude man of his autonomy and value in the ceaseless march of science. Affirming a

stark separation from the naturalist programme provides fodder for those Kantians who seek to sequester Kant's judgement-driven account of action from the scientifically manageable space of causes.

Concerning this familiar image of the intersection of Kant and naturalism, I have noted, without comment, that the contrast is drawn principally on the basis of Kant as a deontological theorist. When we stamp Kant's treatment with the deontological template, Kant not surprisingly becomes a *cause célèbre* for naturalist scorn. It is unsurprising that Dennett suggests that rather than positing 'skyhooks' that miraculously explain human choices by appeal to rational properties of a 'mythical being, the Kantian rational saint', it is better to explain moral behaviour by adopting 'the evolutionary approach, sneaking up on it by gradual increments . . . passing from blind selfishness through pseudo-altruism to quasi-altruism . . . to something that may be quite good enough for all of us'.³

As I have suggested, the familiar description directs the challenge posed by naturalism squarely at the deontological target. It is worth remembering, thankfully, that Kant did not end his contributions to moral theory with the introductory account of obligation found in the *Groundwork*.

There are many elements in the picture beyond the foundation-of-value account of the *Groundwork* that one might invoke to more accurately gauge the relationship with naturalism. Impressive work by Kant scholars over the past two decades has done much to propel Kant well past the narrow and sterile path encouraged by an exclusive focus on deontology. Barbara Herman's work on moral deliberation and Allen Wood's emphasis on the many social features embedded in normative judgments offer but two of many reformist projects that aim to liberate the Kantian approach from charges of being insensitive to context, of being ignorant of the salience of sensuous affects in the cultivation of character, and of discounting the role of culture and history in shaping the conditions in which choice and moral progress are situated.⁴

No doubt revisionist treatments along the lines hinted at here provide a wider basis for articulating the relation between Kant and naturalism. One could develop the case from this point of departure, looking to the elements described as adjuncts to the otherwise naked deontological thrust of the orthodox model.

Nonetheless, I will argue here that Kant's relation to naturalism is brought into sharper relief if one locates the points of contact within the arc of highest good. The struggle to make oneself worthy of happiness is *the* fundamental task Kant ascribes to practical reason. It reflects the governing architecture of Kant's mature theory of action, and it continues, of course, to secure a place—the proper place—for Kant's prized treatment of moral obligation.

³ Dennett, *Freedom Evolves*, 217.

⁴ See Barbara Herman, *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993), and Allen Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999).

2. The Highest Good and the Science of Happiness

There is an adage among jewellers that a gem without a setting is just a pretty stone. The point, perhaps a rather self-interested observation, is that the real value of ‘a piece’ is a function not only of the embedded gem, but also of the structure of the setting itself: it shines because it is well placed. When Kant suggests in the *Groundwork* that a duty-determined maxim shines on its own, like a precious jewel, we do well to take the jeweller’s adage to heart. For Kant, the gleam of morally informed choice is realized in the aim of making oneself worthy of happiness. It is the transcendental structure of the highest good that situates moral claims, and reveals their innate value.

The highest good tradition, on the Kantian reading, takes the form of the synthetic unity of virtue and happiness. The twin requirements instantiated in this synthetic unity cut in two directions: a life of happiness absent the demands of virtue would be a life empty of moral status—a setting with no stone—and in the other direction, a life directed at virtue without happiness would be an unrealized life: ‘For, to need happiness, to be also worthy of it, and yet not to participate in it cannot be consistent with the perfect volition of a rational being’ (*CPrR* 5: 110).⁵ The desire for happiness Kant conceives as the natural goal of our deliberation and action.

Of course, in the pairing of virtue and happiness, happiness is the conditioned good, with the demands of virtue operating as the *supremum bonum* within the *bonum consummatum*. Nonetheless, this should not be read as a subordination of the role of happiness in the good life. As Kant remarks: ‘The human being is a being with needs, insofar as he belongs to the sensible world, and to this extent his reason certainly has a commission from the side of his sensibility, which it cannot refuse, to attend to its interest and to form practical maxims with a view to happiness in this life’ (*CPrR* 5: 61).⁶

⁵ Citations to Kant’s works are provided in translation from the Akademie edition: *Critique of Practical Reason* [*CPrR*], *Groundwork of the Metaphysics of Morals* [*G*], and *Metaphysics of Morals* [*MM*] are taken from Immanuel Kant, *Practical Philosophy*, tr. Mary Gregor (Cambridge: Cambridge University Press, 1996); *Lectures on the Philosophical Doctrine of Religion* [*L*], *On the Miscarriage of All Philosophical Trials* [*MPT*], and *Religion within the Boundaries of Mere Reason* [*R*] are taken from Immanuel Kant, *Religion and Rational Theology*, tr. George Giovanni and Allen Wood (Cambridge: Cambridge University Press, 1996); Immanuel Kant, *Critique of the Power of Judgment* [*CJ*], tr. Paul Guyer and Eric Matthews (Cambridge: Cambridge University Press, 2000).

⁶ While stressing the heterogeneity of virtue and happiness, it should be noted that Kant, like Aristotle, recognizes that the realization of virtuous ends can also be a cause of happiness. Particularly in his later works, Kant stresses that the cultivation of character that disposes one to act virtuously will, of itself, create the conditions for pleasure in the discharge of duty. In this sense, Kant’s mature theory of action is, once again, far from the ‘dead weight of duty’ motif promoted by the deontological framework. On the other hand, I believe it to be mistaken to view the role of happiness in the highest good as linked primarily to the conditions of virtue or the exercise of freedom. The positive feedback loop described operates only with the cultivation of character in full stride. Even where virtue supplements the external, contingent, ground of happiness, it never displaces it. For a good description of the related ‘self-rewarding’ approach to morality and happiness see Paul Guyer, ‘Ends of Reason and the Ends of Nature: The Place of Teleology in Kant’s Ethics’, in *Kant’s System of Freedom and Nature* (Oxford: Oxford University Press, 2006), 169–97, in particular 182–6.

The centrality of happiness for the successful employment of practical reason also serves as a reminder, often obscured in the deontological framework, that the activity of willing is bound up with a serious interest in the instantiation of moral deliberation. The segregation of value from conditions of successful action is an all-too-common false path. In this respect, the synthetic unity of virtue and happiness entails the still stronger claim that, absent of the possibility of representing the efficacy of moral deliberation, even the validity of the moral law itself is drawn into question: 'If, therefore, the highest good is impossible . . . then the moral law, which commands us to promote it, must be fantastic and directed to empty imaginary ends and must therefore in itself be false' (*CPrR* 5: 114).⁷

Kant's repeated rebuke of the one-sidedness of the ancient Stoic attempts to situate the good life entirely within the sphere of virtue offers further confirmation of how the transcendental turn shifts attention away from strategies that seek to identify practical reason exclusively with a narrow interest in sequestered conditions of virtue. According to Kant, even untutored experience is sufficient to refute the Stoic's slide from virtue to the experience of happiness:

Self-contentment arises from morality, while happiness depends on physical conditions. No creature has the powers of nature in its control, so as to be able to make them agree with its self-contentment . . . in the present life happiness itself *will hardly be our lot*, and the Stoics probably exaggerated things very much when they believed that in this world virtue is always coupled with being well-pleased. The most infallible witness against this is experience. (*L* 28: 1089–90)

If we liberate, as I think we should, the picture of Kantian practical reason from the narrow *Groundwork*-inspired deontological emphasis, and reorient the account within the abiding structure of the highest good (detailed in the second *Critique*, and elsewhere), it becomes possible to see a common bond linking Kantianism and naturalism on issues related to happiness. Given the ineliminable requirement for conditioned happiness as a necessary and sufficient condition for a complete life, a goal that practical reason has as its sole aim, there are good reasons to see, minimally, common cause with the emerging science-of-happiness approaches in modern naturalism.⁸ In brief, the Kantian analysis should be open to endorsing practical recommendations that look to individual strategies, and public policies, that amplify the ability to engage nature in ways that are demonstrated to maximize success.

⁷ Also see *CPrR* 143.

⁸ The maturation of the 'science of happiness' approach now makes it a foundational component in the field of behavioural economics. Daniel Kahneman's 2002 Nobel Prize was, in part, a recognition of the important place psychological investigation has once again garnered in the fields of economics and social policy. See Kahneman, A. B. Krueger, D. Schkade, N. Schwarz, and A. Stone, 'Toward National Well-Being Accounts', *AEA Papers and Proceedings*, 94/2 (2004), 429–34, and Ed Diener and Martin Seligman, 'Beyond Money: Toward an Economy of Well-Being', *Psychological Science in the Public Interest*, 5 (2004), 1–31, for two landmark contributions to this emerging field. For two recent book-length treatments of public policy implications see J. D. Trout's *The Empathy Gap* (New York: Penguin, 2008) and Richard Layard's *Happiness: Lessons from a New Science* (London: Penguin, 2005).

A potential shoulder-to-shoulder relationship invites the question of the compatibility of Kant's conception of happiness with modern treatments; are they speaking the same language? Although Kant is not always consistent in his description of happiness, for present purposes the description offered in the *Critique of Practical Reason* is a good general guide: 'Happiness is the state of a rational being in the world in the whole of whose existence everything goes according to his wish and will, and rests, therefore, on the harmony of nature with his whole end as well as with the essential determining ground of his will' (*CPrR* 5: 125).

Linked to this account is the idea that, while happiness is an objective description of a universally felt need, it does not offer an objective practical principle for action. Given the idiosyncratic character of 'wish and will', what makes me happy may well leave you cold. Moreover, what made me happy today may fail to satisfy my desires tomorrow. As Kant puts the point, happiness 'is still only the general name for subjective determining grounds, and it determines nothing specific about it . . . That is to say, in what each has to put his happiness comes down to the particular feeling of pleasure and displeasure in each and, even within one and the same subject, to needs that differ as this feeling changes' (*CPrR* 5: 25). While Kant laments the fact, he concludes that the indeterminacy of the concept of happiness is such that 'the problem of determining surely and universally which action would promote the happiness of a rational being is completely insoluble' (*G* 4: 418).

Regarding this last aspect of Kant's treatment of happiness, a contemporary Kantian might only fault Kant in being perhaps unduly pessimistic about the possibilities of effectively identifying and cultivating the empirical conditions for happiness. When Kant emphasizes the changeable character of happiness, denying it the standing of an objective practical principle, his aim is to highlight the obvious, and important, point that autonomous moral agents with contingent interests will necessarily have divergent conceptions of the good life. However, the stress on the unstable character of happiness can obscure an equally important point. One can rightly reject the thought of a single 'happiness formula' without jettisoning the idea that specifiable empirical conditions can be strongly associated with the satisfaction of individual interests. The obvious candidates include strategies that maximize good health, sufficient resources to meet basic needs, the availability of social interaction, satisfaction in one's employment, etc.⁹

The nascent scientific understanding of happiness, its personal and social conditions, promises much in this area. The burgeoning field of behavioural economics has

⁹ In the *Groundwork*, Kant suggests that even riches, long life, health, and greater cognitive powers offer no necessary relation to happiness: riches creating the conditions for envy and anxiety, long life fostering the possibility of much misery, health leading to bodily excesses that might create greater harm, and enhanced cognitive powers allowing for a clearer eye to the darker reaches of experience (*G* 4: 418). Again, Kant's point is that there is no *necessary* relation between empirical conditions and the desired end. One can grant that point without giving up on the inductive project of modern science to map the most effective footpath towards maximizing individual happiness and social well-being.

matured sufficiently to offer well-tailored practical counsels ranging from macro-economic concerns to the structure of individual choice. For example, at the level of broad economic policy we are now in a better position to appreciate how individual happiness correlates with income level. We know that the declining marginal utility of additional wealth makes social policy that invites large income disparities grossly suboptimal in promoting happiness. Further, there is convincing evidence that a focus on wealth (salary envelope) rather than well-being (family life, social relations, health, etc.), amplifies the negative happiness effects found in relational judgements of well-being (the 'escalator' or 'Jones' effect). We know in the area of social policy, for instance, that proportionally small public investment in preventive health measures can produce substantial effects in avoiding the suffering caused by progressive disease. In the domain of personal decision-making, empirical data strongly support a link between family break-up and the long-term negative impact on the well-being of former spouses and their children. Social policies that make marriage a more informed choice, and the provision for state-funded counselling for marriages at risk, have the potential for substantially improving happiness outcomes.¹⁰

There is an understandable resistance by Kantians to viewing research in these areas as relevant to the analysis of moral life. This shyness may, in part, be motivated by the suspicion that the trajectory of this research belongs properly to a naturalist attempt to frame moral considerations *entirely* within the descriptive sphere. Obviously that meta-project is anathema to any moral theory, Kant's included, that locates the ground of moral value within the prescriptive domain. Nonetheless, Kant's approach is well positioned to appropriate the empirical lessons on offer. As we have seen, the trajectory of Kant's highest good analysis directs serious attention to happiness and its empirical conditions. We are better equipped to lead a life of merited happiness when the empirical conditions for happiness are laid bare.

3. Obligation: Some Kantian Lessons

It may be supposed that the conditioned demand for happiness described in no way touches Kant's treatment of obligation. As virtue and happiness are heterogeneous elements of the highest good, even if one grants that naturalism and Kantianism share a common interest in the conditions that promote happiness, that link is severed at the doorstep of the concept of obligation. As it relates to the inner nature of obligation, this assertion is accurate and will be explored here. Nonetheless, it should be noted that the science-of-happiness contribution is more than a supplement to the natural task of finding happiness for oneself. It also connects immediately to the *obligation* we have to promote the happiness of others.¹¹ With the contemporary push to assimilate the lessons of the *Metaphysics*

¹⁰ These, and many other examples, can be found in Diener and Seligman 'Beyond Money', and in Richard Layard's *Happiness*.

¹¹ The imperfect duty of beneficence requires that we promote the happiness of others. See in particular Kant's discussion in the *Metaphysics of Morals*, Doctrine of Virtue, part II, 6: 448–53. It is important to note

of *Morals* in mind, Kantians should view as friendly the empirical support provided by happiness research as an indispensable aid in fulfilling our imperfect duty to further the happiness of others; effectively realizing our ‘common interest . . . toward those in need’ (*MM* 6: 453). Empirical research on well-being should also prove invaluable in instantiating efficiently the Kantian social obligation (doctrine of right) of material provision (welfare, poverty reduction).¹²

Having acknowledged that happiness issues intrude into the domain of obligation, it is also plain that Kant’s treatment of the nature of obligation constitutes perhaps the most serious point of confrontation with naturalism. The categorical imperative has no obvious naturalistic counterpart. The claim of duty as the principal ground for morally valuable action is out of step, entirely, with the study of the affective influences that inform contemporary approaches to understanding human behaviour. The Kantian demand of universality finds no ready equivalent in, for example, evolutionary accounts of individual or group behaviour.

Sellars’s familiar distinction between the ‘space of reasons’ and the ‘space of causes’ rightfully casts a long shadow over how these divergent accounts are conceptualized. Kant’s appeal to the ‘intelligible character’ of action, his insistence that while movements are individuated by physical descriptions, actions are individuated by maxims, divides cleanly into a ‘two-standpoints’ strategy that highlights the distinctive nomological demands that operate within each sphere.

The stark separation of reasons from causes has not surprisingly provided, as we have seen, the principal target for naturalist attention. In this respect, there is an understandable temptation among Kantians to blunt the force of the separation, and ensuing naturalist critique, by suggesting, for example, a Kantian acceptance of the overdetermination of action by reason and inclination—that reasons *and* causes are the basis for morally valuable action—or by drawing attention to the complex relation between Kant’s account of moral motivation and the moral sense theorists of his day.¹³

that the fulfilment of this duty requires that one promote happiness in accordance with the others’ conception of happiness, hence preserving its agent-driven character.

¹² See Sarah W. Holman’s ‘Kantian Justice and Poverty Relief’, *Kant-Studien*, 95 (2004), 86–106, for an excellent discussion of how a thorough Kantian response might move beyond the demands of virtue (beneficence) to questions of right, grounded in the material conditions necessary for agency.

¹³ Overdetermination strategies of motivation must, of course, wrestle with Kant’s clear separation of inclination and duty; of the passive affective nature of inclination rooted in the space of causes as contrasted with the active intellectual nature of the representation of respect operating in the space of reasons. See Marcia Baron’s *Kantian Ethics Almost without Apology* (Ithaca, NY: Cornell University Press, 1995) for a clear-headed treatment of the subject. Recent attention directed to the powerful influence of Scottish Enlightenment thinkers (Hume, Hutcheson Reid, Shaftesbury) on German philosophy in the 1750–70s has shone a bright light on the historical linkages with Kant, hinting at a less stark separation between Kant and the empiricists of his day (see Manfred Kuehn, Karl Ameriks, and Christine Korsgaard for linkages in the field of epistemology and moral theory). In a recent paper, ‘Naturalist and Transcendental Moments in Kant’s Moral Philosophy’, and ‘Response to Critics’ (*Inquiry*, 50/5, 2007, 444–64 and 497–510) Paul Guyer offers an account of Kant’s maturation as a moral theorist that places substantial emphasis on Kant’s (alleged) early endorsement of a psychological account of the value of freedom, something Guyer sees Kant abandoning later in favour of an unsuccessful transcendental account.

The soft light that these strategies tend to shine on Kant's treatment does little to obviate the underlying worry. The Kantian account of obligation *is* rooted firmly in the space of reasons. Kant's investigation of the transcendental conditions for self-consistent practical judgement is the alpha/omega of his transcendental analysis. Giving this up entails giving Kant up.

A better approach begins with an honest acknowledgement of the bearing of the two-standpoint analysis, with a view to assessing the implications—damaging or not—of the naturalist challenge. A useful start can be made by first drawing attention to the seemingly minor point that Kant's analysis of obligation makes no appeal to non-natural properties. No skyhooks are evident or required.¹⁴ Whatever other problems the naturalist might have with Kant's treatment of obligation, appeals to God's will, the transcendent good, or particular religious traditions are not part of the inventory. Here there is room for agreement.

The axis of confrontation, from the naturalist standpoint, settles rather on the idea that Kant's account of the transcendental conditions necessary for consistent practical judgement fails to engage the 'real' causal structure that underwrites moral phenomena; that the standpoint of action as determined by constraints imposed by self-critical reason use is not a genuine determinant in explaining human behaviour. Explanation and justification for behaviour should be situated, on this reading, in the empirical processes that, bit by bit, account for the 'what' and 'why' of action. At the extreme end, this challenge tends to be mounted in the form of reductive physicalism. More liberal accounts of the naturalist treatment look to the autonomous disciplines of psychology, sociology, anthropology, and economics as mutually supportive frames of reference that provide, when taken together, a suitable empirical treatment of human behaviour.

From the Kantian perspective, the extreme reductivist challenge is hard to distinguish from the genetic fallacy. Imagine, for example, a challenge from the standpoint of neurology to the concept of 'validity' as used in standard truth-functional logic. The neurologist listens politely to the logician's explanation of the property of validity in judgements, and then insists that the 'real' explanation is to be found in the neurological signatures that underwrite such judgements. Now few Kantians would doubt the physicalist claim that moral judgements depend upon brain states for their instantiation. No brain states, no moral judgements. But moving from that dull truth to the sharp edge of reductionism is a leap made comfortably only by those in the grip of the crudest form of eliminativism. In brief, absent such radical a priori commitments, the Kantian is well within her rights to regard this naturalist challenge as a case of changing the subject.

The less radical, but perhaps more far-reaching, naturalist challenge begins with the not implausible thought that moral judgements are just disguised directives to realize

¹⁴ Although no appeals to divine machinations are required in Kant's treatment of obligation, the concepts of God and the Soul, from the practical standpoint, of course do appear in his larger treatment of the highest good. See the following section for details.

one's biologically and culturally grounded conception of living well. In this sense, all imperatives are hypothetical in nature in so far as they provide the means for realizing goals that find their home in biologically rooted, culturally expressed, determinants. Empirical explanation of behaviour takes shape from within the unified plane of each of the social scientific approaches, resulting in a collective unity that is sufficient for purposes of explanation and prediction. From this perspective, the Kantian project of sequestering the putative transcendental grounds of action from the multi-disciplined context of explanation violates, in a deep way, the methodological requirement of unity. Rather than taking physicalism with reductionism as the basis for the challenge, here the more liberal naturalist objection looks to the demand of unity in scientific explanation. Evolutionary, psychological, and anthropological forms of explanation have it individually and collectively; transcendental analysis lacks it.

A Kantian might disarm this worry by first noting that transcendental analysis in no way speaks against understanding behaviour in terms of causal processes, past history, or evolutionary strategies. A generous spirit towards the trajectory of contemporary human sciences should be unproblematic on the Kantian approach. As Kant conceives it, transcendental analysis draws our attention to the inherent twofold character of our engagement with nature: one part the causal structure of behaviour as discerned by the best science of the day, one part a description of agency understood in terms of the demands of valid practical judgement. Explanans and explanandum operate *within* each domain, the demand of unity successfully framing the plane(s) of causal explanation, and providing the nomological framework for valid normative judgement. From the Kantian perspective, the unity objection conflates the appropriate demand for unity within linked domains (the law-determined empirical domains of the relevant empirical sciences), for a (purported but absent) unity between the causal and reason-giving domains. For better or worse, Kant offers a motivated argument for separating the two.

The burden of the argument rests with the naturalist. The task, so conceived, is not simply to explain how obligation manifests itself in social interaction. No doubt the disciplines of evolutionary psychology and anthropology are well positioned to offer the requisite accounts of how *homo sapiens* gradually came to adopt behaviours that frame private obligation and patterns of social sanction. The Kantian is entirely open to this causal account. The Kantian analysis aims at something different. The notion of a categorical imperative, and the requirement of universality, are elements underwriting the modal character (necessity) found uniquely in judgements of obligation. It is *this* explanandum, after all, that is the target of Kant's transcendental inquiry into the nature of obligation: the question of how a categorical imperative is possible. However well-tailored the psychological/evolutionary accounts might be, they seem, again, to offer the wrong type of strategy for excluding the bearing of the transcendental analysis: as wrong-headed as attempting to reduce the patterns of valid inference in mathematical judgements to the psychological/evolutionary propensities that explain the emergence of culturally determined number systems.

In both senses, reductivist and liberal, it is not at all evident how the banishing of the transcendental analysis of obligation is to be mounted, even if we sneak up on it, bit by bit. Strategically, the lack of any discernible nomological linkage speaks against the prospect of discharging naturalism's burden.

4. The Outer Edge of Practical Reason

Modelling practical reason under the rubric of the highest good tradition, in addition to exorcizing the 'rational saint' caricature that bewitches many naturalist readings of Kant, throws into relief an outcome not often associated with Kant: an attentiveness to the enduring presence of tragedy in human affairs.

To see this, it is necessary to recall that, since virtue and happiness are heterogeneous elements, and because happiness is rooted in the contingent conditions of external nature, human flourishing on the Kantian model remains eternally hostage to the vagaries of human experience. In this respect, like modern moral luck theorists, Kant is attentive to the bearing of circumstance and contingency in our struggle to make ourselves worthy of happiness, highlighting the fact that:

it is not always within our power to provide ourselves with happiness, and the course of nature does not of itself conform to merit . . . [that our] good fortune in life (our welfare in general) depends, rather, on circumstances that are far from all being in our control. (*MM* 6: 482)

As nature stands outside our control—prompting our capacity to incorporate its deliverances from outside the space of reason—a triumphalist reading of Kant's account of the good life is wide of the mark in both content and tone.¹⁵ For the Kantian, nature is coldly indifferent to the unity of virtue and happiness:

there is not the least ground in the moral law for a necessary connection between the morality and proportionate happiness of a being belonging to the world as part of it and hence dependent upon it, who for that reason cannot by his will be a cause of this nature and, as far as his happiness is concerned, cannot by his own powers make it harmonize thoroughly with his practical principles. (*CPrR* 5: 124)

As we have seen in the previous discussion of the highest good, without the ability to conceptualize a coordination between deliberation and its effects in the world—that a necessary 'ought' is also a necessary 'can'—practical judgement is set against itself: marooned between the unconditioned demands of practical reason and the indifference of nature. The rejection of the self-sufficiency of the Stoic moral sage, as we saw, serves as a reminder of the vulnerability of the good life to the capricious deliverances of nature.¹⁶ Against this backdrop, without adequate conceptual grounds for

¹⁵ Bernard Williams's triumphalist description of the Kantian project is emblematic of this (mistaken) view. See *Moral Luck* (Cambridge: Cambridge University Press, 1982).

¹⁶ The sharp edge of this dependence and vulnerability is absent in accounts of the highest good that over-emphasize the happiness that derives from self-determined virtuous action. See n. 6.

securing the possibility of representing the realization of the highest good, Kant contends that practical reason would find itself in an *ad absurdum practicum*: the disabling existential state in which practical reason is unable to affirm a condition necessary for its own employment.¹⁷

Kant's way forward, much derided by critics and champions alike, looks to the concepts of God and the Soul. In brief, Kant claims that, on practical grounds, we are entitled to invoke the concept of God as a means, for practical purposes only, to represent the possible union of morality and nature:

Accordingly, the existence of a cause of all nature, distinct from nature, which contains the ground of this connection, namely of the exact correspondence of happiness with morality, is also *postulated* . . . Therefore the highest good in the world is possible only insofar as a supreme cause of nature having a causality in keeping with a moral disposition is assumed . . . Therefore the supreme cause of nature, insofar as it must be presupposed for the highest good, is a being that is the cause of nature by *understanding* and *will* (hence its author), that is, **God**. (CPrR 5: 125)

Following on this, we are permitted, again only for practical purposes, to posit continued existence (soul):

Morality in itself constitutes a system, but happiness does not, except insofar as it is distributed precisely in accordance with morality. This, however, is possible only in the intelligible world, under a wise author and regent. Reason sees itself as compelled either to assume such a thing, together with life in such a world, which we must regard as a future one, or else to regard the moral laws as empty figments of the brain, since without that presupposition their necessary success, which the same reason connects with them, would have to disappear. (CPrR A811/B839)

The truncated character of Kant's justification for the legitimacy of employing these concepts in his resolution of the antinomy of practical reason (in the *Critique of Practical Reason*) has led many readers to question both the cogency and motivations of Kant's seemingly backdoor theistic strategy.¹⁸ Whatever the merits of that challenge, Kant himself consistently adopts this interpretative approach; anticipating it in embryonic form in the first *Critique*, carrying it to term in the *Religion*.

¹⁷ In *Kant's Moral Religion* (Ithaca, NY: Cornell University Press, 1970), 29–30, Allen Wood offers a penetrating analysis of the force of the *ad absurdum practicum* argument.

¹⁸ Lewis White Beck's influential *Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960) no doubt set the tone for near fifty years of questioning concerning the bearing of the practical postulates as they relate to the highest good. Guyer's strategy is to associate the practical postulates' function in the highest good with Kant's moral psychology, arguing that the postulates 'are products of human psychology', best understood as the basis for the cultivation of our sensibility in conformity with the demands of reason, rather than in terms of Kant's own self-declared religious orientation for the postulates. (See Guyer, 'From a Practical Point of View: Kant's Conception of a Postulate of Pure Practical Reason', *Kant on Freedom, Law and Happiness* (Cambridge: Cambridge University Press, 2000), 336.) In a related challenge, other critics have emphasized how the highest good might be conceptually sustained by looking to the historical/social character of human progress. Andrew Reath's 'Two Conceptions of the Highest Good', *Journal of the History of Philosophy*, 26 (1988), 593–619, provides an emblematic statement of this approach.

For the immediate purpose of tracking Kant's sensitivity to the place of tragedy, what is important is not the intramural connection Kant asserts between morality and religion, but rather the implications found at the intersection of the relation of the deliverances of untutored nature and the transcendental permissibility of hope. It begins with a recognition that the transcendental turn, far from being insensitive to the grim empirical setting in which we struggle to make ourselves worthy of happiness, in fact directs our attention to the capricious character of that struggle.

In this respect, it is perhaps too easy for twenty-first-century readers to ignore the shadow of the historical setting in which mid-eighteenth-century Enlightenment thinkers reflected on issues of practical reason. To see Kant's remarks in relevant historical context, it is worth recalling the impact of the single most shattering (both materially and psychologically) event of eighteenth-century Europe: the Lisbon earthquake of 1755. On a sunny Sunday morning, with most inhabitants of the area in the city for Mass on All Saints Day, a powerful earthquake, followed quickly by fire and then a tidal wave, killed 90,000 people, destroying one of Europe's largest cities within the space of a few hours. Beyond the sheer scale of the event, the details that trickled out at the time painted a bleak picture of trapped and badly burned men, women, and children having their screams extinguished by the rush of the sea.

The horror of this event reverberated throughout the intellectual circles of early Enlightenment Europe. Reactions varied greatly, from some priests on the ground viewing the event as divine punishment for sin, to early scientific accounts—a young Kant included—looking to explain the event with reference to the disinterested laws of nature.¹⁹ The correspondence between Rousseau and Voltaire on the implications of the earthquake is perhaps best known. Unlike Rousseau who saw the disaster in terms of human culpability (Rousseau's standard list of the follies of civilized living—density of population, desire for wealth, etc.), Voltaire viewed the event as ultimately beyond recovery. *Candide*, best known among philosophers as a running critique of Leibniz's 'best of all possible worlds' theodicy, finds its satirical voice in Voltaire's rejection of the possibility of assigning any redemptive impulse to the Lisbon tragedy.

Modern naturalist encounters with tragedy, found, for example, in the response to the earthquake and tsunami in Asia in 2004, find no correlate for the theodicy-informed eighteenth century. The contemporary response entirely looks to disinterested nature as the basis for the significance of the event. In the case of the Asian example, the principal talking points following the disaster focused on how quickly a system of detectors could be anchored in the Indian Ocean to give advance warning in future cases, as well as concern for the hygienic disposal of bodies, an appreciation of the challenges of reconstruction, and where possible, grief counselling.

¹⁹ For a recent review of the young Kant's reaction to the Lisbon earthquake see Svend Erik Larsen, 'The Lisbon Earthquake and the Scientific Turn in Kant's Philosophy', *European Review*, 14 (2006), 359–67. For a contemporary overview of the event and its influence on Enlightenment thinkers generally, see Nicholas Shrady, *The Last Day: Wrath, Ruin, and Reason in the Great Lisbon Earthquake of 1755* (New York: Penguin, 2008).

The idea that our engagement with nature should (or could) be conceptualized within the space of a moral architecture has no footing in the modern naturalistic framework. The disenchanted character of nature is shot through to the moral subject, recommending nothing beyond silence to mark the event; the shrug of 'bad luck' displacing even the absurdist echo of satire.

Despite the rather cocooned feel induced by twenty-first-century naturalist responses to misfortune, it takes little imagination to see that, in the large and the small, the reality of tragedy directs us to the outermost boundary of practical reason. Tragedy on the scale of major events is a poignant reminder, writ large, of the gap between our 'wish and will' and the deliverances of nature. At the individual level, tragedy stalks us at great cost in the untimely death of parents, children, and friends.

Although Kant rarely speaks directly to the moral experience and implications of tragedy, his treatment of practical reason is in some ways haunted by it. He tends to register this by noting how we rightly judge as unjust that the wicked flourish and the honest suffer; that in the ways of this world, 'there is absolutely no relation according to justice between guilt and punishment'. That life offers a parade of lives 'led with crying injustice and yet happy to the end' (*MPT* 8: 261).

In such cases, Kant's assertion extends beyond the psychological fact that we bristle when the innocent are subject to calamitous events. Few would doubt that, by nature, we are hardwired to disfavour circumstances that frustrate our ability to instantiate our interests. Instead, Kant is directing attention to a property bound up with the unconditioned practical task of making ourselves worthy of happiness. When those deserving of happiness fail to realize it, or when the brutal and corrupt find life well-pleasing, what underwrites our negative affective response is a practical judgement:

As soon as human beings began to reflect on right and wrong . . . the judgment must inevitably have occurred to them that it could not in the end make no difference if a person has conducted himself honestly or falsely, fairly or violently, even if to the end of his life he has found at least no visible reward for his virtues or punishment for his crimes. It is as if they heard an inner voice that things must come out differently . . . hence there must have lain hidden in them the representation, even if obscure, of something they felt themselves obligated to strive for which would not be compatible with such an outcome. (*CJ* 5: 458)

The psychological response, for Kant, is a second order affect that arises in response to practical reason's unconditioned demand to see realized the synthetic unity of virtue and happiness: we *judge* that something is amiss in the relation of virtue and happiness in these cases. The judgement of injustice is a recognition of that moral fact.

This distinction offers some clarity in distinguishing, for example, Kant's response to tragedy from Voltaire's. Voltaire's shift to parody signals his belief that the affect felt in large-scale tragedy, like Lisbon, simply overwhelms any assignable rational significance. A morally authentic response can only look to the absurd in such cases. Satire and parody become the 'reasonable' reply. Kant, on the other hand, can offer a

diagnosis of the transcendental grounds that provoke the feeling prompted by such horror. Here, the transcendental turn is more telling.

To appreciate the tone of Kant's response, it is important to recognize that the priority attached to judgement over affect does not signal a surreptitious return of the moral sage suitably sequestered from the sting of moral experience. On the contrary, Kant's moral exemplar in such cases looks to the biblical story of Job. Kant's focus, not surprisingly, is not the theodicy discussion found in the Book of Job, but rather on Job's integrity in the face of great loss. Kant's rejection of the theodicy offered by Job's 'comforters' is consistent with his general view that all theodicy requires, *a priori*, an acknowledged *understanding* of the coordination of the space of reasons and causes. Unlike his comforters, Job recognizes, on Kant's reading, the impossibility of such knowledge. As with Job's response, the Kantian approach offers no glad tidings on this score. We are permitted to hope for such a coordination. We are entitled to find strength in the belief that such a coordination is not impossible. However, such faith and hope is ultimately rooted in conditions that allow for the move from practical judgement to action, not from knowledge to practical inferences. This distinction is crucial.

Kant's denial of such knowledge serves two aims. First, as we have seen, it grounds his critique of the possibility of theodicy. Theodicy involves knowledge claims that we are simply not in a position to make. The few priests on the scene who claimed they saw divine punishment in the Lisbon earthquake—apart from lacking in prudential judgement for their own well-being—inferred, as Job's comforters did, that the resulting suffering could only be squared with God's nature on the condition that it was deserved suffering for (some hidden) sin. The Kantian rejection of such inferences is based on his blanket denial that knowledge claims operate in the practical sphere.

The second service it performs is negative: securing the flank of practical reason in its internal demand for a representable coordination between practical judgement and the realization of its ends. In denying the bearing of the theoretical employment of reason, Kant is at pains to stress that this secures the boundary conditions for practical reason's recommendation for rational hope. Hence the lesson from Job, as Kant sees it, lies on the dispositional side of Job's stance: 'only sincerity of heart and not distinction of insight; honesty in openly admitting one's doubts; repugnance to pretending conviction where one feels none . . . these are the attributes which, in the person of Job, have decided the preeminence of the honest man' (*MPT* 8: 266–7).

This is not a description of moral triumphalism in the face of adversity. Neither does it recommend a retreat into irrationalism. Nor does it ignore the crying sense of injustice. Kant sees in Job a resolve to stand firm in a setting that bewilders. Job refuses to engage in a falsification of his moral experience; either in the direction of assuming divine punishment for unknown moral lapses or in endorsing an absurdist orientation in the face of the inscrutable gap between his welfare and his worthiness to be happy. Neither, one might add, does Job deny the empirical reality that he is deeply wronged. In this way, Job's struggle, existential and unceasing, comes to symbolize the demands

of practical reason in a being that must play out his life authentically within the contingency of what the world metes out.

Kant's relation to naturalism, as I have suggested, is thus complex. His rejection of theodicy puts him squarely on the same page as contemporary naturalists. As explanation is rooted in the space of causes, the problem theodicy attempts to solve does not emerge on either view. Although the Kantian will reject the reasons offered by the naturalist for discarding theodicy, there is no light between Kant and naturalism in terms of the result. Both deny its legitimacy.

Where these positions diverge is over the significance of tragedy in human experience. While Kant denies theodicy, he remains appropriately attentive to the empirical reality of tragedy. It is hard to see how naturalism can have any workable concept of the tragic. The disenchanted character of nature envelops the moral subject. This is the calling card of modern naturalism; that human behaviour is of a piece with other patterns of activity in nature. A naturalist might register regret that a life is cut short or, more generally, that human flourishing is rarely coordinate with the deliverances of nature, but a judgement concerning the lack of justice in such engagements with nature is considered senseless.

A brave new world that effectively writes tragedy off the page may be appealing to some. But the burden here, as before, lies entirely with the naturalist. When we are forced to confront the pressing features of our experience—usually in times of crisis or great change—the experience of moral tragedy is rarely far from the centre. That most honest of teachers, great literature, points us in the same direction. Shakespeare's most bleak tragedy, *King Lear*, offers audiences stark images of man's confrontation with disenchanted nature; from monstrous daughters who attack like emerging cuckoos, to a bastard son who outmanoeuvres a loving brother and father, to a King whose concourse with others is rarely more than nature's own currency of reflected power: 'Who loves me the most.' The closing act, where the evil and good fall to the same grave, is wrenching even by Shakespearean standards. The final scene of *Lear* cradling the dead Cordelia, an inversion of the redemptive analogue of the Pietà, signifies the dimensions of the struggle for significance in a world scarred by tragedy.

In the *Critique of Judgment*, Kant offers a similarly stark description of the empirical condition of man (directed in this case to the moral man devoid of faith—Spinoza):

Deceit, violence, and envy will always be rife around him, although he himself is honest, peaceable, and benevolent; and the other righteous men that he meets in the world, no matter how deserving they may be of happiness, will be subjected by nature, which takes no heed of such deserts, to all the evils of want, disease, and untimely death, just as are the other animals on the earth. And so it will continue to be until one wide grave engulfs them all—just and unjust, there is no distinction in the grave—and hurls them back into the abyss of the aimless chaos of matter from which they were taken—they that were able to believe themselves the final end of creation. (*CJ* 5: 452)

It is perhaps not surprising that it is in the Kantian text most focused on the issue of purposefulness that we should find such a desolate description of the human predicament in relation to the significance of the empirical gap between the unconditioned demand to make ourselves worthy of happiness and the contingent deliverances of external nature. Absent the practical postulates, moral experience, on Kant's view, reduces to an *absurdum practicum*: our aims and purposes falling untimely into the one wide grave of chaotic matter. With the postulates in play, practical reason avoids absurdity. It recommends no ordained victories or knowledge of an underlying unity. This is not an escape from, nor an anaesthetizing of, the horrors of experience. The transcendental turn provides merely the scaffolding for representing how deliberation *can be* efficacious given the brute externality of disenchanted nature. Unrehearsed struggle is our solitary commission.²⁰

²⁰ I would like to thank Anna Wilks and Eric Wilson for their helpful comments.

4

The ‘Synthetic-Genetic Method’ of Transcendental Philosophy

Kantian Questions/Fichteian Answers

Daniel Breazeale

In the *Prolegomena*, Kant contrasts the ‘analytic’ method of the latter with the very different method of philosophizing employed in the *Critique of Pure Reason*, which as he puts it ‘absolutely had to be composed according to the *synthetic method*.¹ He further explains that, in the first *Critique*,

I worked on this question *synthetically*, namely by inquiring within pure reason itself, and seeking to determine within this source both the elements and the laws of its pure use, according to principles. This work is difficult and requires a resolute reader to think himself little by little into a system that takes no foundation as given except reason itself and that therefore tries to develop cognition out of its original seeds without relying on any fact whatever.²

In contrast, the ‘analytic’ method of the *Prolegomena* simply assumes ‘that some pure synthetic cognition *a priori* is actual and given, namely *pure mathematics* and *pure natural science*’, in which case ‘we do not need to ask whether [synthetic cognition *a priori*] is possible (for it is actual), but only: *how is it possible*, in order to be able to derive, from the principle of the possibility of the given cognition, the possibility of all other synthetic cognition *a priori*.’³

The *Prolegomena* thus tells us what the proper method of a genuinely transcendental philosophy *is not*. As for what it *is*, we are provided only with a bare *name* (the true method of transcendental philosophy is ‘synthetic’), a specific *locus of inquiry* (transcendental philosophy investigates pure reason itself), and an *admonitory warning* that not everyone may be capable of philosophizing in accordance with the synthetic

¹ The abbreviations used in the footnotes are explained in a list at the end of this chapter. *P*, AA, 4: 263.

² *P*, AA, 4: 274.

³ *P*, AA, 4: 275.

method. But anyone who wishes to learn *how* to philosophize in this manner must turn elsewhere for assistance.

If one then turns to the ‘Transcendental Doctrine of Method’ at the end of the first *Critique* for assistance with this problem, one finds that Kant is here too primarily concerned to explain what the synthetic method of transcendental philosophy *is not*, and far less forthcoming in providing a positive characterization of the same. In the section titled ‘The Discipline of Pure Reason In Its Dogmatic Use’, the reader learns that philosophy, like mathematics, is to be understood as a form of *a priori rational cognition*. But whereas mathematical cognition is grounded upon the construction of its concepts by exhibiting *a priori* the pure intuitions corresponding to the same, philosophy is quite unable to construct its concepts in any similar way, inasmuch as it has no corresponding access to a universally valid realm of non-empirical intuitions. Instead of being able to consider or catch sight of (*betrachten*) ‘the universal in the particular’, philosophical cognition ‘considers the particular only in the universal.’⁴ Nevertheless, both philosophers and mathematicians deserve to be called *Vernunftkünstler* or ‘expert reasoners’;⁵ inasmuch as both are concerned with obtaining synthetic cognition *a priori*. It follows that, though conceptual analysis may be a valuable philosophical tool, it simply *cannot* be the primary method of transcendental philosophizing, for, as Kant writes, ‘at issue here are not analytic propositions, which can be generated through mere analysis of concepts . . . , but synthetic ones, and indeed ones that are to be cognized *a priori*.’⁶

But if the philosopher is unable to resort to *a priori* construction in intuition, then how is he supposed to obtain his synthetic cognitions *a priori*? Alas, on this crucial point, this section is not particularly helpful. ‘To philosophize’, Kant tells us, means ‘to reflect discursively,’⁷ and ‘there is, to be sure, a transcendental synthesis from concepts alone, with which in turn only the philosopher can succeed, but which never concerns more than a thing in general, with regard to the conditions under which its perception could belong to possible experience.’⁸ Philosophy thus provides us with ‘synthetic cognition’ of the rules for synthesizing what is given through empirical intuition, but not with any *a priori* intuition of a ‘real object.’⁹ This description of the task of transcendental philosophy, however, still leaves unanswered the underlying questions concerning Kant’s own philosophical method, namely: *How* does the philosopher arrive at ‘synthetic cognition’—not of experience itself—but of the conditions necessary for the possibility of experience? What *warrants* the specific claims put forward by the transcendental investigator of these necessary conditions? What is the *evidence* for such claims? How can they be *demonstrated*?

If, writes Kant, ‘I am given the transcendental concept of a reality, substance, force, etc., it designates neither an empirical nor a pure intuition, but only the synthesis of empirical intuitions (which thus cannot be given *a priori*), and since the synthesis cannot proceed

⁴ KrV, A713–14/B741–2.

⁵ KrV, A717/B745.

⁶ KrV, A718/B746.

⁷ KrV, A718/B746. Translation modified. Kant writes: ‘Ich würde also umsonst über die Triangel philosophieren, d.i., diskursiv nachdenken . . .’

⁸ KrV, A718–19/B746–7.

⁹ KrV, A720/B748.

a priori to the intuition that corresponds to it, no determinative synthetic propositions but only a principle of the synthesis of possible empirical intuitions can arise from it.' From this Kant concludes that 'a transcendental proposition is therefore a synthetic rational cognition in accordance with mere concepts (*ein synthetisches Vernunftbegriffniß nach bloßen Begriffen*), and thus discursive, since, through it, all synthetic unity of empirical cognition first becomes possible, but no intuition is given by it *a priori*.'¹⁰ Here again, Kant's point appears to be primarily a *negative* one: namely, that philosophy cannot begin with definitions, since its fundamental concepts must simply be 'given' to it.¹¹ But what he fails to explain is precisely how the 'transcendental concepts' in question are supposed to be 'originally given to' or discovered by the philosopher.

Nor is an answer to this question to be found in the section of the Doctrine of Method devoted to the 'Discipline of Pure Reason in regard to its Proofs.' Here too Kant is more concerned to tell us what philosophical proofs *are not* rather than to explain what they *are*. We already know that transcendental propositions are neither *mathematica* nor *dogmata*, inasmuch as they are demonstrable neither by construction of concepts in pure intuition nor by direct inference from concepts.¹² The synthetic method of philosophical proof must therefore be *indirect*, in the sense that philosophy must 'deduce' its synthetic propositions by referring its *previously given* transcendental concepts to some 'third thing': namely, the concept of 'possible experience'.¹³

Useful as such a description of the distinctive character of transcendental propositions and principles (*Sätze und Grundsätze*) may be, it still does not explain *how* the philosopher is actually supposed to 'prove' or to 'deduce' the ostensible synthetic connection between pure concepts that is expressed in his transcendental principles. Nor does Kant here provide his readers with any insight into the precise manner in which mere 'reference' to the concept of possible experience is supposed to function as the 'clue' (*Leitfaden*) or 'guideline' (*Richtshnur*)¹⁴ for a transcendental deduction of the principle in question. Instead, he rebuffs such questions with the following blunt declaration: 'Of the special method of transcendental philosophy, however, nothing can here be said, since we are concerned only with a critique of the circumstances of our faculty—whether we can build at all, and how high we can carry our building with the material that we have (the pure *a priori* concepts).'¹⁵

Thus, in his discussion of the 'proofs of pure reason', Kant is content to repeat his earlier account of the differences between mathematical and philosophical proofs and to remind his readers of the importance of 'possible experience' for the possibility of

¹⁰ KrV, A722/B750.

¹¹ See KrV, A729–30/B757–8.

¹² KrV, A736/B764.

¹³ This claim does not actually conflict with Kant's claim, later in the 'Transcendental Doctrine of Method', that philosophical proofs must be 'direct' rather than 'indirect'. What is rejected in this section is not the legitimacy of referring to some 'third thing' (namely, possible experience) as a 'clue' for proving a synthetic philosophical principle, but rather, the appropriateness within philosophy of 'apagogic' proofs, which purport to prove a proposition by establishing the falsehood of its opposite. A philosophical proof, in contrast, must always be direct or 'ostensive'. 'The direct or extensive proof is, in all kinds of cognition, that which is combined with the conviction of truth and simultaneously with insight into its sources' (KrV, A789/B817).

¹⁴ KrV, A782–3/B810–11.

¹⁵ KrV, A738/B766.

synthetic, 'transcendental cognition'. The only positive characterization Kant provides of his own synthetic method in this section is that 'each must conduct his affair by means of a legitimate proof through the transcendental deduction of its grounds of proof, i.e., directly, so that one can see what claims regarding reason (*Vernunftansprüche*) have to say for themselves.'¹⁶ But as for the kind of evidence appropriate to such a direct or 'ostensive' philosophical proof, Kant once again remains silent; and he maintains a similar silence elsewhere, for instance, in the Jäsche *Logic*, where he characterizes the synthetic method as a 'progressive' one, which 'goes from principles to consequences or from the simple to the composite,'¹⁷ but offers no hint concerning how the transcendental philosopher is supposed to make such 'progress' in applying his own synthetic method.

To summarize Kant's claims—or, better, perhaps, his promises: transcendental philosophy requires a distinctive method of inquiry and proof. Like the method of mathematics, the method of philosophy is *synthetic*, *progressive*, and *a priori*, and it must therefore be carefully distinguished from the synthetic *a posteriori* method of the empirical sciences and from the analytic *a priori* methods of the purely formal sciences. Neither psychology nor logic can serve as methodological models for the philosopher. Nor can mathematics provide a model for the philosopher, inasmuch as philosophy is unable to construct its concepts in intuition. The philosopher must *begin* with concepts acquired from elsewhere and, by relating these to the concept of possible experience, *progress* by means of 'ostensive' proofs to the point where he has established the synthetic *a priori* principles of experience as such. At this point, however, a puzzled reader of Kant may well wonder *what is left for the philosopher to do* once he has dispensed with the methods and tools of direct and indirect logical inference, conceptual analysis, and construction in pure intuition.

Some hint of an answer to this last question may perhaps be contained in Kant's passing use (A719/B747) of the term 'discursive reflection' as a synonym for 'philosophy,'¹⁸ which certainly suggests, as do various other remarks by Kant, that the proper organ of philosophy is neither pure intuition nor scientific understanding, but reason itself in its distinctive, reflective capacity to apprehend its own nature. The question, then, is *how* does reason apprehend itself? As we have seen, Kant's own remarks on philosophical method fail to shed much light on this question; so perhaps the best way to discover Kant's actual method of proceeding in his three *Critiques* is to pay less attention to his own, rather unhelpful methodological pronouncements and more attention to his actual *practice* as a transcendental investigator of theoretical and practical reason and of the power of judgement. Inspired by the work of George Agich, Dieter Henrich, and Huston Smith,¹⁹ I have elsewhere made an attempt to do just this

¹⁶ KrV, A794/B822.

¹⁷ AA, 9: 149. Kant, *Lectures on Logic*, tr. Michael Young (Cambridge: Cambridge University Press, 1992), 639.

¹⁸ See KrV, A719/B747.

¹⁹ George J. Agich, 'L. W. Beck's *Proposal of Meta-Critique and the Critique of Judgment*', *Kant-Studien*, 74 (1983), 261–70; Dieter Henrich, 'Kant's Notion of a Deduction and the Methodological Background of the First Critique', in *Kant's Transcendental Deductions*, ed. Eckart Förster (Stanford, CA: Stanford University

and have proposed an interpretation of Kant's philosophical method that stresses the critical role therein of 'transcendental reflection'.²⁰

Here, however, I am interested in how other philosophers, immediately following Kant, took up the challenge of discovering and expounding the distinctive method of transcendental philosophy, even if this meant that they sometimes found themselves disagreeing with some of Kant's own explicit statements concerning transcendental method. This full story would involve a discussion of K. L. Reinhold, J. S. Beck, and Salomon Maimon, as well as J. G. Fichte. But in this chapter I will limit myself to the latter, and indeed, to the writings of his early or 'Jena' period—a period that actually began in Zurich in 1793 and ended in Berlin in 1800. As we shall see, Fichte's attempts to work out for himself what is involved in the 'synthetic' or 'progressive' method of philosophizing eventually lead him to articulate a conception of transcendental philosophy's proper method as one of *a priori genetic description*, constituting, as he called it, a 'pragmatic history of the human mind'. Fichte's methodological innovations exercised, in turn, a decisive influence upon his own followers, including Schelling, Hegel, and Schopenhauer, but this too is a story that will have to await another occasion. Here I will focus exclusively on the 'genetic method' of the early *Wissenschaftslehre*. Though Fichte's relationship to Kant has been frequently studied,²¹ little attention has been paid to the question of philosophical method; yet this is one of the more illuminating points of comparison between the *Wissenschaftslehre* and the Critical philosophy, inasmuch as Fichte clearly believed that his own explicit and repeated attention to methodological issues represented one of his more important *advances* upon Kant, who, he complained, 'appears to have philosophized far too little about his own philosophizing'.²² As he wrote to a friend at the end of 1793, 'I am deeply convinced that Kant merely pointed to the truth, but without exhibiting it or proving it. . . . No one has yet understood him—those who think they have understood him best have understood him least. Only by reaching Kant's results in one's own way will one have understood Kant'.²³

He therefore made every effort to follow a rigorous and perspicuous method in his philosophizing and to *explain* and *elucidate* the same in his so-called 'critical' (or as we would say, 'metaphilosophical') writings and remarks. He wanted his

Press, 1989), 29–46; Houston Smith, 'The Role of Reflection in Kant's *Critique of Pure Reason*', *Pacific Philosophical Quarterly*, 80 (1999), 203–23.

²⁰ 'Kant's Critical Philosophy: In Search of a Method' (plenary address at the Bicentennial Kant Conference, Auckland, New Zealand, 11 July 2004).

²¹ For a recent and particularly thorough example of such a study see Armin G. Wildfeuer, *Praktische Vernunft und System: Entwicklungsgeschichtliche Untersuchungen zur ursprünglichen Kant-Rezeption Johann Gottliebe Fichtes* (Stuttgart-Bad Cannstatt: Frommann-Holzbock, 1999).

²² Fichte to Reinhold, 4 July, 1797, *GA* iii/3, 69; *EPW* 420.

²³ Fichte to F. I. Niethammer, 6 Dec. 1793 (*GA* iii/2, 20–1; *EPW* 368). See too Fichte's letter to Heinrich Stephani, mid-December, 1793 (*GA* iii/2, 28; *EPW* 371), where he remarks: 'Kant's philosophy, as such, is correct—but only in its results and not in its reasons. This singular thinker looks more marvelous to me every day. I believe that he possesses a genius that reveals the truth to him without showing him why it is true. I believe that in a few short years we shall have a philosophy which will be just as self-evident as geometry.'

readers to understand as clearly as possible how every single claim advanced within the *Wissenschaftslehre* is related to and completely determined by those that precede it, for only then can they follow what he described to Schiller as ‘my synthetic ascent.’²⁴ Even before arriving in Jena, as he was preparing for his first public presentation of his new system before an informal audience in Zurich, Fichte declared that, ‘My presentation is always synthetic; I never simply throw out my thoughts in the form in which I may have thought them in my study; but rather, I think them, discover them, and develop them before the eyes of my listeners and along with them.’²⁵

Following Kant, Fichte sought to distinguish the properly ‘synthetic’ method of transcendental philosophy from the competing methods of logical inference, conceptual analysis, and empirical proof. But unlike Kant, he took care to provide his readers with several, elaborately detailed, positive accounts of what he meant by the ‘synthetic method of philosophizing,’ and he made a great point of explicitly following this method—or, as we shall soon see, these *methods*—of synthetic deduction in all of his systematic treatises. Though I shall be making reference in what follows to all of Fichte’s major works from the Jena period—including the *Foundations of the Entire Wissenschaftslehre* (1794/95), the lectures on the *Foundations of Transcendental Philosophy (Wissenschaftslehre) nova methodo* (1796/99), and the *System of Ethics* (1798)—I shall begin by focusing upon some features of the *Foundations of Natural Right* (1796/97), a work that the author himself explicitly recommended for the particularly rigorous manner in which it observes and illustrates the ‘synthetic method’ of philosophizing.²⁶

The *Foundations of Natural Right* opens with a provocative series of claims. The *Wissenschaftslehre* is to be sharply distinguished from virtually all preceding philosophies, inasmuch as it is grounded in a clear insight into the uniquely reflexive structure of *Ichheit* or ‘I-hood’ itself, in which subject and object, ideality and reality, are originally and synthetically united in the sheer and spontaneous self-assertion of the I as an I. The task of transcendental philosophy is to provide a complete description of all of the *additional* acts of the I that are necessary for the possibility of—and thus implicit in—the originally postulated self-positing of reason (which, for Fichte, is simply another name for the I itself), and the method appropriate for such a *descriptive* task is that of meticulously attentive *self-observation* or *inner intuition*. Fichte characterizes the resulting description of the activity of consciousness as ‘*genetic*’, inasmuch as each of the described acts is explicitly recognized by the philosophical observer of the same to be the condition for the possibility of the previously described acts, as well as a condition for the acts that follow—all of which are thus conditions for the possibility of the absolutely and originally posited unity of self-consciousness itself.²⁷

²⁴ Fichte to Schiller, 27 June 1795, *GA* iii/2. 358; *EPW* 394 (translation corrected).

²⁵ Fichte to J. K. Lavater, beginning of Feb. 1794, *GA* iii/2. 60.

²⁶ See *SS*, *GA* i/5. 104–5.

²⁷ One of the passages in which Fichte most clearly indicates this reciprocal relationship between ‘what is conditioned’ and ‘what conditions’ in the *Wissenschaftslehre* occurs in §1 of *System der Sittenlehre*, *GA*, i/5. 41; *The System of Ethics*, tr. Daniel Breazeale and Günter Zöller (Cambridge: Cambridge University Press, 2006), 29,

Since such a genetic description of the necessary acts of the I claims to demonstrate that the I, in order to posit itself as an efficaciously acting free self, must also posit a 'world' with a certain necessary structure, transcendental philosophy also purports to provide an a priori, genetic account of the world of human experience.

The 'strict method' of proof to which Fichte pledges to adhere in his treatise on natural right, as in all of his other transcendental treatises of this period, is a fundamentally *descriptive* one, which, as such, presupposes on the part of the transcendental philosopher a radically purified and rigorously disciplined power of 'inner observation' or 'intellectual intuition'. What distinguishes such a method from the casual introspection of the armchair psychologist, according to Fichte, is that the object of the former is not the contingent 'facts of consciousness', but rather the *necessary* and *universal*, and hence a priori, conditions of the same.²⁸ These conditions, in turn, are identified with the *necessary* and *pure* acts of the self-positing I—the very acts that are identified and related to one another in a genetic account of the constitution of self and world. But how, it will be asked, is an empirically real philosopher able to identify such extraordinary mental acts, which never appear, as such, within ordinary empirical self-consciousness?

In order to place himself in a position to observe such acts and thereby to employ the 'genetic' or 'synthetic method' of philosophizing, the would-be philosopher must begin by elevating himself, by means of a freely undertaken act of global 'abstraction', from the standpoint and concerns of ordinary life—including everything that pertains to him as a distinct individual—and then carefully 'attend to' (*aufmerken*) or 'reflect upon' what remains after he has abstracted from everything from which he is at all *able* to abstract. What will then remain as the object of philosophical observation or reflection, Fichte assures us, will be pure self-consciousness itself, a spontaneously initiated and initially empty self-reverting activity or *Tathandlung*. But he certainly does not ask his reader to take his word for this; on the contrary, he insists that if the student or reader cannot and does not perform this initial act of 'abstraction' for himself then he can save himself the (considerable) trouble of proceeding any deeper into the fabled intricacies of the *Wissenschaftslehre*.

The ability to accomplish this first task of abstraction, by means of which alone one can elevate oneself from the 'ordinary' to the 'philosophical' standpoint, is not something that 'comes naturally', at least not to most of us, but is instead a skill

where he writes that the proposition, 'I necessarily find myself', refers to a kind of self-consciousness that is 'exhibited in the foundational portion of the entire *Wissenschaftslehre*—not, to be sure, as a matter of fact, for as such it is something immediate, but rather in connection with all the other types of consciousness, as reciprocally conditioned and conditioned by the latter'.

²⁸ Fichte's rejection of what would today be called 'psychologism' begins with his criticism of Platner and continues in his famous criticism of the effort of Schmid and others to base philosophy upon an appeal to the 'facts of consciousness'. It becomes even more explicit and pointed in his writings of 1799–1801. See e.g. 'Antwortschreiben an Herrn Professor Reinhold' (1801) and 'Aus einem Privatschreiben' (1800), where Fichte notes that, unlike psychology, which deals simply with the 'facts of consciousness', the *Wissenschaftslehre* deals with 'what one finds to be the case only when one discovers oneself' (GA i/6. 387n.; IWL 174n.).

that has to be deliberately cultivated by determined effort and practice.²⁹ Hence the *Wissenschaftslehre* (as Fichte clearly realized only in 1795, that is, only *after* the publication of Part I of *Foundations of the Entire Wissenschaftslehre*, with its misleading talk about 'single first principles' and its appeal to logical propositions, such as that of identity) begins not with a self-evident first principle nor axiom, but instead with a 'summons' or 'invitation' (*Aufforderung*) to the would-be philosopher, with a 'challenge', if you will, 'simply to think the I' (in the case of the foundational portion of the system, as presented in the lectures on *Wissenschaftslehre nova method*)—or 'to think the I as an individual' (in the case of the *Foundations of Natural Right*) or 'to think the I as possessing a causal power of its own' (in the case of the *System of Ethics*).

By accepting and responding successfully to this summons, one will at the same time succeed in raising oneself from the 'ordinary' to the 'philosophical' standpoint, from which standpoint one will, Fichte assures his readers, be able to detach the necessary acts of the I—necessary, that is, for the possibility of the originally posited *Tathandlung* through which the I is present to itself *as* an I— from the 'products' of the same—that is, from the world and objects of ordinary experience, including the empirical I. One will thereby obtain cognitive access to an entirely new realm of mental activity and will be able to observe the acts in question and to describe them in their necessary, sequentially ordered relation to one another (that is, 'genetically'). Thus, the transcendental philosopher must not only make the effort required to think these 'necessary thoughts', that is, he must not only actually perform all the mental acts necessary in order to 'think the I' (which means, according to Fichte, that he must think in accordance with the fundamental law of all thinking, the so-called law of reflective opposition or principle of determinability), but he must also reflect upon his own thinking of the same, in order to see what *else* he has to think in order to 'think the I'.

Since the universal criterion of 'reality', according to Fichte, is the feeling of being compelled to present (*darstellen*) something in a specific way (and hence in the presence to consciousness of 'representations accompanied by a feeling of necessity'³⁰), and since the philosophical observer finds himself similarly 'compelled' to record the self-constitutive acts of the pure I in the manner and in the precise order in which he both re-enacts and observes them, he must also attribute a certain 'reality' to the same—though a reality very different in kind from that of sensible objects or empirical states of (outer or inner) consciousness. Fichte's point is that the *Wissenschaftslehre* does not *fabricate* its objects of reflection (the necessary acts of the I), but instead,

²⁹ Fichte sometimes refers to the skill in question as 'philosophical genius' or "spirit" in the distinctively philosophical sense, but he always tends to treat this not as an innate capacity, the possession of which distinguishes some human beings from others, but as a power implicit in spirit or mind as such and hence as a talent that anyone might acquire through determination and practice. On this topic see GA ii/3. 316–17. EPW 199–208; BWL GA i/2. 143; EWP 127–8; and GNR, GA i/3. 326; FNR 6.

³⁰ This is the language employed in the *Wissenschaftslehre nova methodo* and in the two well-known 'Introductions' of 1797, which are based upon these same lectures. See e.g. GA i/4. 186–7; IWL 7–9.

discovers and *observes* them, which is why it deserves the name *Reellephilosophie*.³¹ As he puts it in his lectures on *Wissenschaftslehre nova methodo*:

The *Wissenschaftslehre* itself does not generate any new cognition. It merely observes the human mind in its original generation of all cognition. . . . Ordinary, commonsense understanding, on the other hand, does something and observes only the product of its own doing. It fails to notice that through its own doing it also generates time, etc. The *Wissenschaftslehre* pays attention to this 'doing' itself. This is the synthesis we have been discussing, and the *Wissenschaftslehre* has to display this synthesis independently of analysis, . . . Only in this way does it obtain a genetic understanding of the origin of our representations.³²

Though this process can certainly be described as an 'analysis' of the I, the cognitions pertaining to such a *Reellephilosophie* are themselves *synthetic* or *ampliative*. They are not derived from an analysis of the *concept* of the I nor by logical inference from any preceding set of concepts and principles. Instead, as we have seen, these transcendental cognitions are grounded upon actual (inner) intuitions of a series of pure and necessary acts of the I. The obvious parallels between this method of philosophical self-observation and description and Husserl's method of phenomenological reduction and eidetic intuition suggest that Fichte's method of philosophizing, at least as characterized so far, might well be described as 'phenomenological'.³³ The Fichtean philosopher is an 'observer of a real act of thinking accomplished by his own mind'; and hence, if he has properly accomplished the initially required acts of abstraction and reflection, an observer of 'reason in its *original and necessary* way of preceding, through which the philosopher's I and everything that is for it exist'.³⁴

As previously noted, the *Wissenschaftslehre* does not purport to be a collection of mere 'facts of consciousness'. On the contrary, an absolutely essential and distinctive feature of the same, without which, according to Fichte, it could not hope to accomplish its task of providing a transcendental derivation of ordinary consciousness from the concept and intuition of the I's original self-positing, is that this philosophy describes an interrelated *series* of actions, along with the *products* of the same. This is a point Fichte stressed in comparing his own philosophy to the desiccated Kantianism of his colleague C. E. Schmid:

If one philosophizes in Professor Schmid's manner, the object about which one philosophizes is always something *static and fixed*. In the *Wissenschaftslehre* the object is something *active*, and it is *presented in its activity*. This latter science does not propose to justify any system of things, but rather to describe a series of acts. It allows the I to act before its eyes, while it observes this acting. The I which it observes is not what might be called the 'philosophizing I', which, as always happens when one contemplates anything, is lost in contemplation; instead, it is the common,

³¹ See sections I and II of the Introduction to *GNR*, as well as §7 of *BWL*.

³² *WLnM[K]*, GA iv/3. 480; *FTP* 380.

³³ See Breazeale, 'Fichte's *Nova Methodo Phenomenologica*: On the Methodological Role of "Intellectual Intuition" in the Later Jena *Wissenschaftslehre*', *Revue Internationale de Philosophie*, 4 (1998), 587–616.

³⁴ *GNR*, GA, i/3. 316 and 316n.; *FNR* 7 and 7n. (translation modified).

ordinary I. All descriptions which occur in the *Wissenschaftslehre* are, therefore, genetic descriptions. Thus the curtain which keeps the unconsecrated from entering the *Wissenschaftslehre* veils the entire domain of this science as well, for those who are unable to produce within themselves the act we have described and who do not actually produce this act really do see nothing and obtain nothing.

It is only in this way that the *Wissenschaftslehre* is able to accomplish what was demanded of philosophy. We have, for example, been told frequently enough which predicates describe representation. But what we wanted to know was what *the act of representing* (*das Vorstellen*) really is. And this is something that can be presented only genetically, that is, insofar as one's own mind is engaged in this very act of representing.³⁵

This is why Fichte's synthetic method of philosophizing must be characterized not only as 'phenomenological', in the sense indicated, but also as 'genetic'. And it is for precisely this reason that Fichte himself sometimes described his own philosophy as a 'history'—and indeed, a 'pragmatic history'—of the human mind.³⁶ What makes such a 'history' *pragmatic* lies precisely and quite specifically in the fact that what it describes is a *process*, a necessary series of interconnected acts by means of which something—namely consciousness and experience—is *produced*. In other words, a pragmatic history of the mind is, by definition, a *genetic description* of the same. (Fichte obtained this concept of genetic history from Ernst Platner's *Philosophical Aphorisms*, the work he used as a textbook in this 'Introductory' lectures of 'Logic and Metaphysics'.³⁷)

The philosopher must, insists Fichte, describe each determinate mental action he observes 'with respect to its form.' That is to say, he must describe how the act in question is determined by those acts that precede it and determines those that follow it. It is this 'genetic' method of derivation that produces the systematic 'form' of philosophy itself. But the philosopher also has to describe 'what emerges for reflection in this acting [i.e. the products of the same, including our ordinary representations of the physical world]. By doing this, he simultaneously provides proof of the concept's necessity, determines the concept itself, and shows its applicability.'³⁸ Since Fichte understood this 'genetic' method to be the sole viable method of transcendental philosophizing,

³⁵ GA i/3. 256–7; EPW 325–6.

³⁶ Fichte famously describes the *Wissenschaftslehre* as a 'pragmatic history of the human mind' (GWL, GA i/2. 364–5 and BWL, GA i/2. 147) or simply as 'the history of the human mind' (GA ii/3. 107) or as 'the history of consciousness as it comes to be (*die Geschichte des entstehenden Bewußtseins*)' (WLnm[K], GA iv/3. 464).

³⁷ A 'pragmatic history of the human mind' is a *genetic* account of self-constitution of the I in the form of an ordered *description* of the various acts of thinking that are presupposed by the act of thinking the I. See Fichte's note in his Platner lectures (GA ii/4. 51–2): 'Now, however, Platner establishes the following . . . [:] What is it? It is a pragmatic history, and something of this sort is what I want to lay down [*verzeichnen*—though from directly opposed first principles [viz., not from mere "facts of consciousness"]—"pragmatic" = how it comes into being. "History" is fiction, indicates the genetic manner of the presentation (*pragmatisch wie es zu Stande kommt. —Geschichte, ist Fiktion, giebt den genetischen Gang des Vortrags*). For a detailed examination of this topic, see ch. 4 of Breazeale, *Thinking through the Wissenschaftslehre: Themes from Fichte's Early Philosophy* (Oxford: Oxford University Press, 2013), 70–95.

³⁸ GNR, GA i/3. 319; GNR 9 (translation modified).

he naturally concluded that this must have been Kant's actual method as well, despite Kant's failure to make this clear.³⁹

There are, of course, numerous difficulties and problems implicit in such a conception of the proper method of transcendental philosophy, and chief among these is the issue of the alleged 'purity'—hence the 'universality' and 'necessity'—of the descriptions in question. Fichte was by no means insensitive to this problem, even if his proposed solution is not without problems of its own. As I have already noted, he apparently believed that the purity of the philosopher's inner intuitions and hence the universality of his descriptions is, so to speak, *guaranteed* by the *completeness* of the initial act of free abstraction which precedes his series of self-observations.⁴⁰ But the only way he could imagine to 'demonstrate' such a tendentious claim was simply to issue a *challenge* to his readers and critics to *perform for themselves* the requisite acts of abstraction and inner observation postulated by the *Wissenschaftslehre*, so that they could then verify its propositions for themselves. In this sense, Fichte always insisted that the *Wissenschaftslehre* must be understood as an *experimental* enterprise, which every inquirer has to test for himself by conducting the requisite 'thought experiment', inasmuch as 'one can grasp the truth of this system only by reproducing these self-observations within one's own consciousness'.⁴¹

Not only must the philosophical observer attend to certain *objects* of reflection (namely, certain mental actions and the products of the same), he also has to *produce them*, which is to say, he has to *generate them through thinking*. He therefore is assigned a *dual role* by Fichte's genetic method. On the one hand, he is the 'observer' of those acts that are required for the self-constitution of subjectivity. As such, he is, in Fichte's words, a passive 'spectator of the theater of his own mind'.⁴² On the other hand, he must also take the stage and assume the leading role as the impersonator of the pure I or of pure reason itself as it finds that in order to posit itself at all it must posit itself as an embodied individual in a spatio-temporal world of material objects—which he is able to do only because, in his free resolve

³⁹ 'The aim of Kant's writings was to introduce such a philosophy and to do away with all merely formal philosophizing' (GNR, GA i/3. 317; FNR 7).

⁴⁰ See Breazeale, 'The "Standpoint of Life" and "The Standpoint of Philosophy" in the Jena *Wissenschaftslehre*', in Albert Mues (ed.), *Transzendentalphilosophie als System: Die Auseinandersetzung zwischen 1794 und 1806* (Hamburg: Felix Meiner, 1989), 81–104.

⁴¹ 'The philosopher is not a mere observer; instead, he conducts experiments with the nature of consciousness itself and turns to himself for answers to his specific question. This is a system for persons who are able to think for themselves. It cannot be grasped merely by reading and studying. Every person must produce it within himself, particularly since no fixed terminology will be introduced. . . . One can grasp the truth of this system only by reproducing these self-observations within one's own consciousness' (WLnM[K], GA iv/3. 339 and WLnM[H], GA iv/2. 25; FTP 101–2). Fichte explained this to his own students as follows: 'In this course we will be conducting experiments. I.e., we will compel reason to provide us with answers to specific, systematically calculated questions; then, for the purpose of science and as an aid to memory, we will formulate the results of our experiments in concepts' (WLnM[K], GA iv/3. 350; FTP 120).

⁴² *Eigne Meditationen über ElementarPhilosophie/Practische Philosophie*, GA ii/3. 70.

to philosophize, he has deliberately *abstracted* from all that pertains to him as an individual I.⁴³

To be sure, the genetic method we have been describing is not the only philosophical method employed by Fichte. Like any other philosopher, the transcendental philosopher has to take proper care to clarify or to 'explicate' (*erörtern*)⁴⁴ what is contained in or 'pertains to'⁴⁵ the various concepts with which he is dealing, no matter where or how he may have obtained them. This means that he must analyse his previously established concepts and draw inferences (*Folgerungen, Schlüsse, Corrollaria*) from his previously established propositions or theorems,⁴⁶ and, where appropriate, incorporate the latter into formal, syllogistic arguments.⁴⁷ This is an essential part of any philosopher's toolkit. In addition to this, however, Fichte's Jena presentations of his system frequently employ yet another type of thinking, this time as a *means of discovery*. Over and over again in these writings one encounters a distinctively 'synthetic' kind of *thinking*, which can only be described as 'dialectical' in character. Fichte first displayed his talent for this kind of thinking in Part II of the *Foundations of the Entire Wissenschaftslehre*, in which the original 'contradiction' between the limited I and the limited Not-I serves as the motor that drives the philosophical inquirer to posit ever higher 'syntheses', in an ultimately vain effort to resolve the contradiction in question, but it is also a prominent methodological feature of the *Foundations of Natural Right*.⁴⁸

⁴³ Whereas the 'formulaic philosopher' is conscious only of his own thinking (and hence only of the acts of his own individual I), the 'real (*reelle*) philosopher', by virtue of the act of abstraction with which he begins, is able to become the observer of the necessary operation of 'reason as such', according to its own inner laws (GNR, GA i/3. 316n.; FNR 7n. (translation modified)).

⁴⁴ Thus e.g. after having established, in the second theorem of GNR, that a finite being cannot ascribe to itself free efficacy in the sensible world without ascribing the same to others and thus presupposing their existence, Fichte describes the next task as that of *analysing* this proposition, in order to *explicate* (*erläutern*) and *clarify* what has already been established (GNR, GA i/3. 344; FNR 34).

⁴⁵ GNR, GA i/3. 390; FNR 87 (translation modified). See too GNR, GA i/3. 384; FNR 79 (translation modified), where Fichte announces that 'our path now leads us to an explication (*Erörterung*) of the *inner* conditions of such reciprocal interaction'. As Fichte explains in *BWL*, to 'explicate' a concept scientifically is 'to assign it a place' in relation to other concepts (GA i/2. 127; EPW 113–14).

⁴⁶ GNR, GA i/3. 344; FNR 34.

⁴⁷ See e.g. Fichte's comment, in §7 of GNR, that his proof (*Erweis*) of the external conditions for the possibility of a community of free being 'is based solely on the presupposition of such a community, which is itself grounded on the possibility of self-consciousness', and that 'all the conclusions up to this point have therefore been derived (*abgeleitet*), by way of mediate inference (*mittelbare Schlüsse*), from the postulate "I am I"' (GNR, GA i/3. 384; FNR 79).

⁴⁸ In §8 of the *Foundations of Natural Right* e.g. Fichte remarks that one can trust another person to subject himself to the law of right only if one knows that the other will not in the future violate one's own rights; and one can be sure that the other person will actually respect one's rights only if one believes him to have sincerely subjected himself to the law of right. In this case, concludes Fichte, 'That which is grounded is not possible without the ground; and the ground is not possible without that which is grounded. We are thus caught in a circle. We shall soon see how, in such a case, one must proceed in accordance with the synthetic method' (GNR, GA i/3. 394; FNR 91). Faced with such an apparently 'insoluble' contradiction, one must engage in an act of *creative* or *imaginative* thinking. 'In order to eliminate the contradiction, these two elements will be synthetically united in accordance with the method demonstrated in the *Wissenschaftslehre*' (GNR, GA i/3. 395; FNR 92). That is, we must introduce a new thought, that of a 'third party' who will act as the future guarantor of the present compact between the first two parties. In this manner, an apparently analytic 'contradiction' provides the occasion for a new *synthetic* inference.

Such dialectical-synthetic thinking proceeds by making explicit a contradiction implicit in a previously derived set of propositions and then actively 'seeking out' (*aufsuchen*) some new, 'higher principle' that might allow one to avoid the objectionable contradiction and is *therefore* 'necessary'. Unlike conceptual analysis, logical inference, or syllogistic reasoning, this 'dialectical' method of derivation is thoroughly *synthetic*, in the sense that the new principle that resolves or abolishes (*hebt, aufhebt*) the contradiction in question is not already 'contained in' and thus cannot be analytically derived from the problematic set of concepts and propositions that it is supposed to synthesize. Furthermore, since it is not derived from experience, but is instead a product of pure thinking, this new principle is also 'a priori' and thus represents a synthetic a priori extension of our cognition.

Arguments of this sort are encountered over and over again in the *Foundations of Natural Right*, in book ii of the *Foundations of the Entire Wissenschaftslehre*, and elsewhere, and on occasion are explicitly formulated in terms of a 'thesis' and its 'antithesis', which produce a contradiction, which is in turn *aufgehoben* by a new 'synthetic' proposition.⁴⁹ Fichte sometimes even gives a quasi-mathematical appearance to such arguments by formulating them like problems in geometry: Problem (*Aufgabe*): to seek a certain concept = X that will resolve a preceding contradiction 'in accordance with a strict method'.⁵⁰

An essential feature of this kind of thinking is that it can be guided by no *algorithm*. At each step, one encounters new *contradictions* that cannot be analytically resolved, fresh *problems* to be solved, and novel *challenges* to be met. To say, with Fichte, that an appropriate solution to such problems is something we have to 'seek out' is to recognize that every such problem must be dealt with on its own terms and that each requires a fresh exercise in *creative problem-solving*. Neither past nor present experience can offer us any guidance in such cases, for here we remain within the realm of pure reason⁵¹—or, if one prefers, the 'space of reasons'—and we must therefore seek our solution via

For some additional examples of Fichte's dialectical/synthetic mode of proof, see GNR, GA i/3. 407–8, 447–8, 453, and 456–7; GA i/4. 12, 59–60, and 159–60. The famous 'deduction', in §§5 and 6, of the human body and of the two organs of the same—as well as the corresponding deduction of the various external 'media' through which these organs can be affected—also contains many examples of 'dialectical' inference of the sort described.

Consider too the case of *money*, the necessity (and determinate character) of which Fichte 'derives' from the contradiction between (1) the civil guarantee that once a citizen has fulfilled his duties of protection and support with regard to the state he has an absolute, unrestricted right to the remainder of his possessions and (2) the state's absolute right to appropriate the products of each citizen's labour in order to make possible the exchange of goods, thereby allowing everyone to live off his own labour. The right to property thus seems to be contradicted by one of its immediate implications. Here again we are forced to 'seek out' a solution and thereby to demonstrate the *necessary* introduction by the state of legal tender, which allows the state to control the substance of its citizens' possessions, but not the form thereof (GNR, GA i/4. 41).

⁴⁹ GNR, GA i/3. 397; FNR 94–5. For other explicit 'thesis/antithesis' formulations of the synthetic method of proof, see GNR, GA i/4: 41–2 and 59–60; FNR 206–7 and 226–7.

⁵⁰ GNR, GA i/3. 433; FNR 134.

⁵¹ See e.g. Fichte's claim to have deduced the necessity of the ephorate, not from any appeal to experience, but 'from pure reason' (GNR, GA i/3. 449n.).

imaginative thinking. This, presumably, is why Fichte says, in a 'remark on synthetic method' in his lectures on *Wissenschaftslehre nova methodo*, that this dialectic method of deduction is the 'most difficult method of all' and attributes the failure of his original 1794/95 presentation of the foundations of his system to be understood to the fact that the latter made such heavy use of the dialectical-synthetic method.⁵²

Anyone who has ever tried to construct a proof in geometry, to solve a mystery, or indeed to demonstrate a philosophical proposition, is surely acquainted with the kind of 'capacity' presupposed by Fichte's dialectical-synthetic method. But such a person will also be acquainted with another, problematic feature of all such 'solutions' or 'proofs': namely, the problem of establishing the *necessity* and *uniqueness* of the proposed 'synthetic principle'. It is surely one thing to construct or to imagine a 'workable' solution to a problem or an ingenious way of avoiding a theoretical contradiction by introducing a new, synthetic principle, but it is something else altogether to claim that the solution in question is the *only possible one*. And this is why, for of all of his brilliance and talent as a dialectician, Fichte did not, in the end, rest the case for his philosophy on the same and appealed instead to results obtained by the direct or 'genetic' or 'synthetic' method of demonstration.

Fichte's clearest discussion of the relationship between these two different philosophical methods comes near the end of Part II of the *Foundations of the Entire Wissenschaftslehre*, where he explains the difference between what he calls an 'empty formulaic philosophy' and a 'system of real thinking', such as the *Wissenschaftslehre*, and then relates this distinction to his own, very different methods of philosophizing in the remainder of Part II and in most of Part III of this same work. Hitherto, he notes, we have been considering mere 'possibilities of thought' brought forth by the spontaneity of our own mind; but now, with the postulation of the reality both of the *Anstoß*—the infamous 'check', a term that is replaced in the *Wissenschaftslehre nova methodo* by the more perspicuous term 'feeling', as designating the original and philosophically incomprehensible limits of the finite I—and of the creative power of the I (here described as the power of productive imagination), our inquiry acquires an altogether firmer footing, not in mere abstract thought possibilities, but in a distinctive sort of *fact*: the pure fact of consciousness itself. (This pure fact is by no means to be confused with the empirical 'facts of consciousness'.) And, with this, we make the transition from the general method of synthetic-dialectical *thinking* to the distinctive *genetic-phenomenological method* of transcendental philosophizing. From now on the system becomes 'a system of real thinking' or 'pragmatic history of the human mind'. Here, declares Fichte, at the 1794/95 *Foundations of the Entire Wissenschaftslehre*, we have arrived at a critical 'turning point for philosophical reflection'; and thus, 'from this point on, the entire reflection will stand on a completely different level and will possess an utterly different meaning'. Whereas previously we reflected upon objects

⁵² *WLnm[H]*, GA iv/2, 107–8; *FTP* 248–9.

of thought (namely, synthetic propositions) that were produced by our own creative, dialectical thinking, ‘in the future series of reflections we will be reflecting only upon *facta*: the object of this reflection is itself a reflection; namely, the reflection of the human mind, in accordance with the basic law of all reflection (that is, the law of reflective opposition) upon the datum indicated therein’—namely, the *Anstoß*, or its original feelings of its own limits, as well as its awareness of its own free efficacy—‘Hence, in the future series of reflections, the object upon which we are reflecting will not itself first be *produced* through this same act of reflection, but will merely be *raised to consciousness*.’ As the philosopher is now able to observe for himself, this object of consciousness is a product of the necessary act of reflection (or self-positing) of the I that the philosopher is observing.⁵³

Another way to explain the relationship between the two methods is to note, with Fichte himself, that Part II of the *Foundations* can provide us with no more than an *apagogic* or indirect proof of its conclusion: it demonstrates that unless we are prepared to accept the demand for absolute causality, the identity of the I will have to be given up. But the *reality* of this demand for causality, writes Fichte, requires something more: It ‘must also be capable of a direct and *genetic* demonstration; it must make itself believable not only by appeal to higher principles, which, without this demand, would be contradictory, but it must actually be capable of being *deduced* from these higher principles themselves, so that one can obtain insight into (*einsehen*) how such a demand arises in the human mind.’⁵⁴ Fichte reiterated this same point some years later in §10 of Part II of the *System of Ethics* with respect to the postulated ‘drive to obtain self-consciousness’: ‘First of all,’ he notes, ‘we have to derive the drive in question. In the preceding section we proved that if there is no such drive, then the self-consciousness of I-hood is impossible; for in that case consciousness of an indeterminacy, which is a condition for the self-consciousness of I-hood, is impossible. That constituted an indirect proof [that we possess] such a drive. A direct—that is to say, a genetic—proof, from the concept of the I itself, must now be undertaken.’⁵⁵

The kind of ‘proof’ or ‘deduction’ to which Fichte is here referring is not a logical argument based upon inferences from higher principles nor a dialectical argument of the type considered above, but is instead a ‘genetic deduction’⁵⁶ of the type described earlier, the type embodied in his own ‘pragmatic history of the human mind’. Over

⁵³ GWL GA i/2. 363–5; SK 197–9 (translation modified). This contrast between the two methods of philosophy is further elaborated in the first section of the introduction to GNR, ‘How a Real Philosophical Science is Distinguished from a Merely Formulaic Philosophy’ (GA i/3. 313–18; FNR 3–8).

⁵⁴ GWL, GA i/2. 404; SK 239 (translation modified). ⁵⁵ SS, GA i/5. 132–3; SE 133.

⁵⁶ GWL, GA i/2. 432; SK 265. See too the key section of part one of the *System of Ethics* titled ‘Genetic Deduction of the Consciousness in Question’ (SS, GA i/5. 47ff. SE 36ff.). The consciousness in question is one’s consciousness of one’s original being, which turns out to be one’s consciousness of one’s own freedom or ‘power of causality by mere concepts’; and the ‘genetic deduction’ of the same is simply a description of the series of those mental acts which are required if one is to posit one’s own activity. See too Fichte’s comments on the connection between ‘deduction’ and ‘my synthetic method,’ in *Aus einem Privatschreiben* (GA i/6. 395–6; IWL 175–6).

and over again in his Jena writings, he insists on this point: the deductions or derivations provided in his system differ from those found in any preceding systems precisely because they are 'genetic explanations' or 'genetic accounts'. And this they *have* to be, because what the *Wissenschaftslehre* purports to describe is something that can be grasped in no other way. 'When one has to explain something for which the words are lacking, one then has to explain the thing itself, i.e., one must explain it genetically. I posit myself, and, in doing this, I pay attention to the fact that I posit myself in a particular manner and that I can posit myself only in this way.'⁵⁷

By employing such a method, the philosopher acquires a kind of knowledge or cognition that contrasts sharply with ordinary 'factual cognition', which appeals simply to 'facts lying within ordinary cognition'. An appeal to the latter, 'unguided by philosophical reflection', says Fichte, 'produces nothing but a deceptive popular philosophy, which is to say, no philosophy at all'.⁵⁸ In contrast, the kind of cognition obtained within transcendental philosophy deserves to be called 'learned' or *genetic cognition*, a distinction that Fichte glosses as follows in his *System of Ethics* (in this case, with respect to our real cognition of our moral nature):

On the other hand, one may refuse to remain content with the immediate perception and, in his thinking, may not stop with the facts, but may demand to know the grounds of what one has perceived. Such a person is not satisfied with factual cognition but demands genetic cognition; he wants to know not merely that such a compulsion is present within him, but he also wants to see how it originates. Were he to obtain the desired cognition, this would be a *learned* (*gelernte*) cognition; and in order to obtain it, he would have to raise himself above the standpoint of ordinary consciousness.⁵⁹

* * *

Let me conclude this brief survey of Fichte's genetic method with two remarks, the first on the relationship between the method of geometry and of philosophy and the second concerning the relationship between the genetic method of philosophizing and the original genesis of consciousness described by this method.

One rather striking implication of Fichte's view of transcendental method is that the differences between the methods of mathematics and philosophy are not as immense as Kant claimed. To be sure, philosophy cannot 'construct' its concepts in pure space or time, but nevertheless it can and must *construct them on the basis of pure (intellectual) intuition*, guided by synthetic thinking and formal logic, but always grounded in immediate reflection upon the necessary acts of the I. This explains why philosophy cannot begin with simple definitions, but must, like geometry (as least as that was understood by Fichte), begin with a *postulate*, with a summons to engage in a certain

⁵⁷ *WLn*m[K], GA iv/4, 337; *FTP* 100.

⁵⁸ *GWL*, GA i/2, 363–4; *SK* 297 (translation modified).

⁵⁹ *SS*, GA i/5, 33–4; *SE* 20.

intellectual activity. Concepts acquired in this manner are ‘genetic concepts.’⁶⁰ Just as the geometer has to ‘construct’ a sphere in order to obtain an accurate concept of the same, so must the transcendental philosopher ‘construct’—or rather, ‘reconstruct’—the I in order to obtain his genetic concept of the same.

Not surprisingly, therefore, Fichte often appealed to his own use of ‘genetic explanations’ in order to characterize what he considered to be one of his most significant advances beyond Kant. Because—and only because—it ‘proceeds genetically’ is the *Wissenschaftslehre* able to answer the question Kant could never answer satisfactorily, or at least not to the satisfaction of his idealist followers: namely, *how do we come to ascribe objective validity to certain representations?*⁶¹ The *Wissenschaftslehre* can accomplish this, boasted Fichte, only because ‘we always proceed genetically, i.e., by self-observation of the “how”.’⁶²

Fichte was at least vaguely aware of the parallels between the methods of the geometer and of the transcendental philosopher from a very early date. Indeed, in his ‘Zurich Lectures’ on the *Wissenschaftslehre*, delivered several months before he arrived in Jena in the spring of 1794, he was already exploiting these parallels. Thus, in the fourth lecture in this series (‘*Wissenschaftslehre* in Relation to Geometry’) he endorses the Kantian concept of ‘demonstration’: ‘For Kant, to *demonstrate* something means (in accordance with the etymology of the word) to make something visible in a sensible representation by dissecting the latter, and one can do this *only* when the concept in question is supported by a sensible presentation, that is, when one can construct it.’⁶³ It follows that the truth of any such demonstration is something that can be grasped only by somehow one actually performs—‘through his own

⁶⁰ ‘A still higher question remains to be answered, namely: *how* can a state begin absolutely, or how can the absolute beginning of a state be thought? In order to answer this question one would have to provide a genetic concept of freedom; one would have to generate this concept before our eyes. We have just now accomplished this. It is not the case that the state that is begun absolutely is simply connected to nothing at all, for a finite rational being necessarily thinks only by means of mediation and connections. The connection in question, however, is not a connection to another being, but to a thinking.

In order to exhibit the concept in this manner, however, one certainly has to follow—and has to be able to follow—the path taken by the *Wissenschaftslehre*; that is to say, one has to abstract from all being as such (from the fact) and commence with what is higher than all being, with intuiting and thinking (with the acting of the intellect as such)’ (SS, GA iv/5, 52; SE 41).

⁶¹ ‘Kant did not answer the question: How is it that we come to ascribe objective validity to certain representations? But the *Wissenschaftslehre* succeeds in answering just this question. We attribute objective validity to a representation whenever we affirm that, in addition to the representation itself, there also exists some thing that corresponds to the representation but is independent of it. What distinguishes the representation from the thing is that I have produced the representation, but I have not produced the thing. The *Wissenschaftslehre* asserts that, when we are dealing with representations that are supposed to be present within us necessarily, we are simply forced to assume that something external corresponds to them; and it demonstrates this genetically’ (WLnM[K], GA iv/4, 328; FTP 84 (translation modified)). ‘And we also obtain, along with this, a genetic understanding of *how* it is that we come to assume that something external to us is given [to us]’ (WLnM[H], GA iv/2, 104; FTP 242).

⁶² GA iv/2, 33; FTP 121.

⁶³ GA iv/3, 33. For Kant’s view of demonstration, see §57 of the *Critique of the Power of Judgment*: ‘concepts of understanding must, as such, be demonstrable at every time; that is to say, the objects that correspond to them must be capable of being given at any time in intuition (pure or empirical)’ (AA 5: 342).

actions'—this construction for himself. Only, in Fichte's words, if 'I, I who am the demonstrating subject, am [also] the subject who engages in the act of constructing',⁶⁴ can I be convinced by such a proof.

After duly taking note of Kant's denial that philosophy is able to construct *its* concepts, Fichte registers his dissent by insisting that philosophy does indeed possess intuitions, albeit non-sensible ones, and hence is in a position to produce construction-based demonstrations of its own. But the actions of construction that have to be performed in the case of the *Wissenschaftslehre* are necessary (and normally unconscious) acts, in contrast to freely undertaken constructions of the geometer, and thus a difficult, second act of reflection is required in the former case in order to be aware that one is even engaging in such a construction (though, as an I, one, *per hypothesis*, always is). Finally, the particular acts performed within the *Wissenschaftslehre* are directed not at something outside the I, but rather, at the I itself, which also makes it difficult to become conscious of one's acting.

But the most important difference between geometry and philosophy noted by Fichte in these lectures is that in geometry one is first conscious of performing certain actions (drawing a circle, say) and then aware of the products of the same. A philosophical *deduction*, on the other hand, *starts* with the 'products' (space, time, the causal structure of the sensible world, for example).

In this case, therefore, what the proof has to show is not *that* something is so, but rather, that it is so by virtue of an action of our mind, or, as *Kant* puts it, that the propositions in question are true *a priori*. In *geometry*, the action is *given*, and what has to be sought is the *product*. In *Wissenschaftslehre*, the *product* is given, and what has to be sought is the action by virtue of which it is there. Geometry asks, *Quid Facti?* *Wissenschaftslehre* asks *Quid Juris?*⁶⁵

⁶⁴ 'Ich der demonstrierende bin der Konstruierende' (GA iv/3. 35).

⁶⁵ GA iv/3. 36. This fascinating passage continues as follows:

Thus, with respect to its inner essence, the method of proof employed in *Wissenschaftslehre* is precisely the opposite of the one employed in geometry, and the external differences between geometry and *Wissenschaftslehre* follow from this. If, therefore, there is in *Wissenschaftslehre* only *one* undoubted fact, the fact that grounds the entire science, and if what must be proven is that all the other facts [established in this science] are indubitable: then the only way to prove this is by showing that even this first fact could not be a fact unless all the others were facts as well.

Just as surely as this first fact is a fact, then the others that follow from it must be facts as well. Their quality, as facts, is derived from the quality of the first one, as a fact. Consequently, *philosophical proofs* are *deductions* (*Deduktionen*), and a deduction is the sort of proof that has here been dissected and explicated; that is to say, a deduction is a proof that proceeds by means of relations of equivalence, as demonstration does as well, but a deduction should not merely show that something is, but should also show that it is a *fact*.

The grounding principle of all demonstration is as follows: because a certain action has occurred within consciousness, an action that, in itself, was not necessary, then consciousness must be modified in a certain manner, if it is to remain a *single unity*.

In contrast, the grounding principle of deduction is as follows: because consciousness is a single unity, then a certain action must necessarily have occurred.

This explication must make it clear that *Wissenschaftslehre* is and must be fully the equal of geometry with respect to the precision of its proofs and that demonstration has no advantage over deduction in this respect.

Despite these important differences between philosophical ‘deductions’ and geometrical ‘demonstrations’, both proceed *genetically*, that is, by means of construction in intuition, though in the case of the latter, this occurs with reference to sensible intuitions and by means of freely undertaken acts of construction, whereas the *Wissenschaftslehre* describes acts necessary for the self-construction of the I and hence of ordinary experience, and it obtains access to the same by means of additional acts of *intellectual* or *inner* self-intuition.

Yet there is some ambivalence in Fichte’s views on this subject; indeed, in a letter written at almost exactly the same time as the Zurich lectures we have just been discussing, we find him *denying* that philosophy can construct concepts in intuition and expressing his agreement with Kant on this point.⁶⁶ But this ambivalence was soon left behind and Fichte pursued his project of constructing a philosophy that would be, in his words, ‘just as self-evident as geometry’.⁶⁷

By end of the Jena period (which actually concluded in Berlin in the year following his departure from Jena in 1799), at a time when he was struggling (unsuccessfully) to revise his lectures on *Wissenschaftslehre nova methodo* for publication under the title ‘New Version of the *Wissenschaftslehre*’,⁶⁸ Fichte was also making his most determined effort to emphasize the parallels between the methods of the geometer and of the philosopher and to provide his system with an external form that closely paralleled that of Euclid’s *Elements*. Thus each section of this ‘New Version’ begins with a ‘postulated’ act of thinking, which allows the philosopher to ‘construct’ within the rational space of ‘pure intuition’ a new result (i.e. a ‘theorem’) and draw conclusions from the same in the form of ‘corollaries’. If, in early 1794, he sometimes seemed to endorse Kant’s claim concerning the inability of philosophy to construct its concepts, by 1800 he was unequivocal in his rejection of Kant’s position.⁶⁹ Philosophy, he now assures his readers, in a public (and, as it turned out, sadly premature) announcement of the ‘New

It is important to remember that this was written at a time when Fichte’s thoughts concerning his new system were still very much in flux. He would soon abandon e.g. all talk about ‘facts’ of consciousness in favour of reference to the ‘acts’ of the same.

⁶⁶ ‘But isn’t it true that philosophy, unlike geometry and mathematics, is quite unable to *construct* its concepts in *intuition*? Yes, this is quite true; it would be unfortunate if philosophy were able to do this, for then we would have no philosophy, but only mathematics. But philosophy can and should employ *thinking* in order to *deduce* its concepts from one single first principle which has to be granted by everyone’ (Fichte to F. V. Reinhard, 15 Jan. 1794, GA iii/2. 40).

⁶⁷ Fichte to J. F. Flatt (draft), Nov. or Dec. 1793 (EPW 366). See the identical claim in Fichte’s mid-Dec. 1793 letter to Stephani (EPW 371).

⁶⁸ *Neue Bearbeitung der W.L. 1800*, GA ii/5. 319–402. For an analysis of this manuscript, focusing upon the concept of construction and stressing the ‘mathematical’ structure of Fichte’s method of demonstration in this fragment, see Breazeale, ‘Die Neue Bearbeitung der *Wissenschaftslehre* (1880): Letzte “frühere” oder erste “spätere” *Wissenschaftslehre*’, *Fichte-Studien*, 17 (2000), 43–67.

⁶⁹ ‘Philosophy, accordingly, would be the cognition of reason itself by means of reason itself—through intuition. The first point was Kant’s important discovery, even though he failed to carry it through to completion. The second point, which expresses the condition for carrying out such a project, was added by the *Wissenschaftslehre*, which, for just this reason, is a completely new science.’ [Ankündigung:] ‘Seit sechs Jahren’ (dated 4 Nov. 1800), GA i/7. 159; IWL 193.

Version of the *Wissenschaftslehre*, is nothing less than ‘the *mathesis* of reason itself’;⁷⁰ and, since it is a ‘kind of mathematics’, it also possesses all of the epistemic advantages of the same, including ‘immediate self-evidence’, ‘thoroughgoing determinacy’ of its results, and ‘the same irrefutability as mathematics’. ‘The *Wissenschaftslehre* . . . simply does not admit the validity of any concept whatsoever that it has not produced within its own boundaries by intuition; and none of its concepts count for it as anything more or other than what is contained within intuition.’ Hence, as in the case of geometry, so too in the case of philosophy: ‘there can be no dispute whatsoever concerning the *Wissenschaftslehre* and no argument against it’;⁷¹ since, after all, such a philosophy is not ultimately based upon ‘argument’ at all.

As for the relationship between the genetic method of transcendental philosophy and the original self-construction or ‘genesis’ of the I: both are ‘synthetic’ in character, and both involve a series of necessary mental acts. Because consciousness itself is a synthetic, self-constructive *process* or *activity*, only a genetic description of the same, based not on arguments and concepts, but rather upon a progressive series of direct intuitions, can provide any real cognition of the same.

Despite such isomorphism, one must not overlook the crucial *differences* between the *real* synthesis of the subject-object and the philosopher’s *artificial* re-enactment of the same. The latter is necessarily discursive, whereas the former is not. The original self-construction of the I occurs, as Fichte puts it, ‘all at once’; it has, so to speak, always already happened, whereas the discrete moments patiently distinguished and genetically connected to one another by means of philosophical reflection and deliberate thinking are, as Fichte emphasized in the *Sunclear Report* (1801), no more than an *artificial reconstruction* of the I’s original, complex act of *self-construction*. For this reason, Fichte’s insistence that the *Wissenschaftslehre* is a system of ‘real philosophical thinking’ must be taken with a certain grain of salt. The ‘pure acts’ this science describes do not really (that is to say, originally) occur as such, in isolation from the other acts, and hence they do not actually occur in the *temporal* order in which they are intuited and described by the transcendental philosopher.

What is really ‘real’ for the early Fichte is the finite, embodied self and its material and human world, along with the moral law, the moral agents, and the moral world-order. In comparison with all of this, the constitutive acts described by transcendental philosophy are mere *fictions*, as Fichte himself conceded on at least a few occasions.⁷² But they are not *arbitrary* ones, as can be verified, according to Fichte, by anyone who takes the trouble to conduct for himself Fichte’s own audacious experiment in ‘thinking the I’, thereby providing himself with a ‘genetic demonstration that—and how—the sort of

⁷⁰ ‘[Ankündigung:] Seit sechs Jahren,’ GA i/7. 160; IWL 195.

⁷¹ ‘[Ankündigung:] Seit sechs Jahren,’ GA i/7. 160–1; IWL 196–7.

⁷² See e.g. GA ii/3. 160–1; i/3. 403–4; iv/1. 197–8; iii/4. 360–1; and i/7. 249–50. For detailed discussion of these passages, see Breazeale, ‘Fichte’s Philosophical Fictions’, in Daniel Breazeale and Tom Rockmore (eds), *Essays on the Later Jena Wissenschaftslehre of J. G. Fichte* (Evanston, IL: Northwestern University Press, 2002).

consciousness with which we are ordinarily familiar flows from our consciousness of ourselves.⁷³

Whether this version of the synthetic-genetic method of philosophizing really is up to the demands that Fichte made on it is, of course, a difficult question and one that plagues not just the *Wissenschaftslehre* but German idealism in general. Though I certainly do not propose to address this question here, I nevertheless hope to have shown that a careful consideration of 'Fichte's genetic method' has much to contribute to one's understanding and assessment of the problematic claim, so central to this entire tradition, that the propositions of transcendental philosophy are 'universal and necessary' (but not analytic), are 'uniquely determined', and possess 'real content' of their own.

Abbreviations

AA	<i>Kants Werke</i> . Akademie-Textausgabe (Berlin: de Gruyter, 1968).
KrV	Kant, <i>Kritik der reinen Vernunft</i> . English translation by Paul Guyer and Allen W. Wood, <i>Critique of Pure Reason</i> (Cambridge: Cambridge University Press, 1998). Page references to the original and second editions (A and B), provided in the English translation as well.
BWL	Fichte, <i>Ueber den Begriff der Wissenschaftslehre</i> (1794).
EPW	Fichte: <i>Early Philosophical Writings</i> , ed. and trans. Daniel Breazeale (Ithaca, NY: Cornell University Press, 1988).
FNR	Fichte, <i>Foundations of Natural Right</i> , ed. Frederick Neuhouser, tr. Michael Baur (Cambridge: Cambridge University Press, 2000).
FTP	<i>Foundations of Transcendental Philosophy (Wissenschaftslehre) nova methodo</i> , ed. and trans. Daniel Breazeale (Ithaca, NY: Cornell University Press, 1992).
GA	J. G. Fichte-Gesamtausgabe der Bayerischen Akademie der Wissenschaften, ed. Reinhard Lauth, Hans Gliwitzky†, and Erich Fuchs (Stuttgart-Bad Cannstatt: Frommann-Holzboog, 1964–).
GNR	Fichte, <i>Grundlage des Naturrechts</i> (1796/97).
GWL	Fichte, <i>Grundlage der gesamten Wissenschaftslehre</i> (1794/95).
IWL	Fichte, <i>Introductions to the Wissenschaftslehre and Other Writings</i> , ed. and tr. Daniel Breazeale (Indianapolis, IN: Hackett, 1994).
P	Kant, <i>Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können</i> . English translation by Gary Hatfield, <i>Prolegomena to any Future Metaphysics that will be able to come forward as a Science</i> , in Kant, <i>Theoretical Philosophy after 1781</i> , ed. Henry Allison and Peter Heath (Cambridge: Cambridge University Press, 2002). Page references to AA, IV, provided in the English translation as well.

⁷³ *WLn*m[K], GA iv/3. 481; *FTP* 381.

- SE Fichte, *System of Ethics*, tr. and ed. Daniel Breazeale and Günter Zöller (Cambridge: Cambridge University Press, 2005).
- SK Fichte, *Foundations of the Entire Science of Knowledge*, tr. Peter Heath, in Fichte, *Science of Knowledge*, ed. and tr. John Lachs and Peter Heath (Cambridge: Cambridge University Press, 1982).
- SS Fichte, *System der Sittenlehre* (1798).
- WLnm[H] Fichte, *Wissenschaftslehre nova methodo* ('Halle Nachschrift', 1796/97).
- WLnm[K] Fichte, *Wissenschaftslehre nova methodo* ('Krause Nachschrift', 1798/99).

5

Fichte's Anti-Skeptical Programme

On the Anti-Skeptical Strategies in Fichte's Presentations of the *Wissenschaftslehre*

1794 to 1801/2

Rolf-Peter Horstmann

The question concerning the substantive philosophical interest, or otherwise, of post-Kantian idealist philosophy, of what has come in short to be known as 'German idealism', has naturally received very different explicit and implicit answers at different times. The entire range of implicit answers is represented in the now enormous and almost unmanageable body of contributions which in one way or another have formulated their own negative or positive responses with regard to the specific theses, theories, and *topoi* of this epoch of German philosophy. These answers can be described as implicit ones in the sense that they are less concerned with clarifying or elucidating the philosophical content of the relevant idealist sources than they are with exploiting the philosophical potential, or indicating the conceptual limits, of these sources in relation to 'new' and 'different' insights of the present. In this connection it has become very common to appeal to the canonic texts of Fichte, Schelling, or Hegel when seeking to support or to contest some contemporary claim or other. The fact that we appeal at all to elements of this idealist tradition as relevant points of reference for our own concerns itself implies a certain position concerning the substantive philosophical value, or otherwise, of the works in question, but precisely a merely implicit one.

As far as explicit answers to our original question are concerned, these are scarcely less numerous than the implicit ones, and have led to an extraordinarily broad spectrum of possible interpretations which then furnish the background for the judgement of philosophical relevance in each case. From the abundant range of these interpretations, some have proved more viable than others, some have appeared more profound and original than others, and a small number have even succeeded in dominating the discussion concerning the merits of German idealism

for a considerable period of time.¹ No one would really wish to claim that the world of competing and mutually supplanting interpretations represents an ordered cosmos of its own, a world that is already organized by fundamental agreement on a common principle of interpretation in confronting the various theories and systems characteristic of this epoch in the history of philosophy. Yet we cannot fail to recognize that over time a certain consensus has emerged concerning the themes and areas that must specifically be addressed in any serious engagement with German idealism, that is, one which hopes to achieve a more or less adequate and appropriate perspective upon this philosophical movement. Amongst these themes we must certainly include the one that is generally labelled, if not informatively characterized, as the problem of 'self-consciousness', the 'I', the 'subject', or 'subjectivity'. That the question of 'self-consciousness' does indeed play a significant role for all the post-Kantian idealist philosophers, and is clearly acknowledged as an absolutely fundamental philosophical issue, is too obvious to require detailed justification here. What is far less obvious, however, is how or why this question came to occupy such a prominent place in the philosophical efforts and investigations of these thinkers. If we wish to pursue the question concerning the ultimate reason for this concern with self-consciousness, the answers we may wish to consider will depend on the perspective from which we identify and define the genuine philosophical interest of the issue of 'self-consciousness'. In the first place, therefore, we shall have to specify more precisely what it ultimately was that interested the idealist thinkers when they interested themselves in the problem of self-consciousness.

According to one influential line of interpretation, their interest was directly connected with the need to clarify the role of first-person consciousness itself, understood as a particular form of object-consciousness, that is, to explore the question how the phenomenon of the 'I', grasped as an object of knowledge, can then be distinguished from other objects of knowledge, or to explain how knowledge of a self is possible at all. Considered from this perspective, self-consciousness becomes a problem of psychology, one that is primarily interested in how I can establish a cognitive relation to myself. Now it is certainly true that this psychological aspect of the problem of 'self-consciousness' was also important for the representative thinkers of German idealism. But a perspective that is basically focused on the psychological problem of self-consciousness can hardly claim to have adequately brought out the specific and systematic relevance of this question for the general overarching theories propounded by the idealist thinkers. And it cannot do so—precisely because a perspective which describes the idealist interest in 'self-consciousness' in this restricted manner fails to

¹ Over the last few decades, it is undoubtedly the work of Dieter Henrich that has furnished the most influential attempt, and this far beyond the limits of the German context, to redefine the specific philosophical character, assess the intellectual achievements, and develop the conceptual perspectives of post-Kantian idealist thought. From his early essays on Fichte up to his most recent contributions, he has impressively pursued the aim of contextualizing the thought of the idealist period in terms of the fraught constellation of problems associated with the themes of 'self-consciousness', 'monism', and 'speculative thinking'.

recognize that this question was never the sole issue of interest for the German idealists, but always played a specific role only within the context of much more comprehensive and far-reaching projects of their own.

One of these projects was principally concerned with establishing a theory of knowing, understood as a theory both about what 'knowing' means and about how knowing is possible. Now it is not difficult to show that it was this project which directly engaged the philosophical efforts of the German idealists, at certain times almost exclusively, but to some extent throughout the entire period, and that it was precisely in the context of this broader project that the question of self-consciousness initially proved so important. We simply need to examine the way in which the idealist thinkers explicitly described—in their various 'Prefaces', 'Introductions', and 'Introductory Observations'—what they themselves took to be the specific object of their philosophical enquiries in each case. And we can immediately see that in the important works which were specifically concerned with presenting the underlying philosophical programme, the writers in question, when attempting to characterize their own respective tasks in a provisional manner, frequently refer to knowing and cognition, the grounding of knowledge, the first principles of knowledge and cognition, etc., but do not present the intrinsic nature or character of self-consciousness as the central point of discussion. This is particularly evident in the case of Fichte and the various early versions of the *Wissenschaftslehre*, but it is equally true of Schelling's essay *Of the I as the Principle of Philosophy* or his *System of Transcendental Idealism*, of Hegel's essay on *Faith and Knowledge*, and even of the first edition of the *Science of Logic*.

It might of course be objected that concentrating simply on such introductory manifestos and proclamations is not the most convincing way of substantiating the suspicion that the psychological interpretation of the question of 'self-consciousness' is an unsuitable point of departure for accurately and appropriately identifying the central philosophical intentions of the post-Kantian idealist thinkers. But there is a further and more important point to bear in mind. For we must consider what it was, as the philosophers concerned frequently conceded, that made the task of clarifying knowledge and the grounds of knowledge appear so urgent to them in the first place. And here it soon becomes quite clear that it was concerns and anxieties concerning the possible plausibility of certain skeptical objections that had been raised, primarily with regard to Kant's theoretical and practical philosophy. For these objections had contributed to the general conviction that any defensible philosophical theory had first of all to present a theory of knowledge which was rightly confident of countering the most resolute and tenacious skeptic. We can properly claim, therefore, that substantiating a programme for the theory of knowledge that was immune to skeptical doubts should be regarded as the principal, and openly defended, purpose of all the systematic philosophical labours of the German idealists, at least insofar as we are speaking of philosophy as an essentially foundational or grounding discipline. This claim, at this point, is merely a general thesis for which the ensuing discussion will have to provide the relevant evidence and justification.

It is important to recognize, therefore, that for Fichte, Schelling, and Hegel alike, the question of 'self-consciousness' is fundamentally connected with the specific context of the theory of knowledge in general. In the case of Fichte and Schelling, this connection was quite direct and explicit in the earlier phase of their thought, while it was rather more indirectly present in their later conceptual formulations (and Hegel also followed them in this regard). In this cognitive-epistemic context the question of self-consciousness did not primarily interest them as a psychological problem at all. It interested Fichte and Schelling above all, at least for a time, because they both believed, for different reasons, that self-consciousness plays a quite specific role with regard to the grounding of knowledge: it seems that we can appeal to self-consciousness as something that is fundamentally immune to all skeptical objections. The specific character of this philosophical interest in self-consciousness reveals an interesting point which almost inevitably escapes our attention if we interpret the central role of self-consciousness in the development of post-Kantian idealist theories as an essentially psychological problem. In short, these thinkers were principally interested in knowledge, rather than in self-knowledge.

If these considerations are indeed correct, then we are naturally prompted to ask the question concerning the precise ways in which the problem of knowledge is addressed, and the consequences that ensue from these approaches, in the context of post-Kantian idealist thought. And this is the question that forms the substantive focus of the following discussion. Our question takes Fichte's philosophy as its immediate historical point of reference, but it will also vouchsafe certain perspectives, at least in passing, on the work of Hegel and Schelling. I shall concentrate on Fichte in particular for three reasons. In the first place, I believe that, amongst all the post-Kantian thinkers that are relevant here, it was principally Fichte who recognized the full complexity of the question of 'knowing' and attempted to do justice to this complexity. In the second place, the example of Fichte is particularly well suited, in my view, to show just how strongly the idealist interest in questions of 'self-consciousness' was actually bound up with fundamental considerations arising from the theory of knowing in general. And in the third place, finally, it seems to me that focusing upon the theme of 'knowing' lends an interconnected character to Fichte's thought as a whole which it certainly lacks when he is interpreted principally as a theorist of self-consciousness. In this manner, as I further hope, we may avoid the rather unfortunate tendency of contrasting an earlier Fichte (up to around 1800) with a later Fichte (after 1800) and thus establishing an all too abrupt, and ultimately unintelligible, opposition between the two.

I

If the attempt to establish a theory of knowledge which can withstand all skeptical objections does indeed represent a central motivation for the systematic philosophical efforts of the post-Kantian idealists, then we must also naturally ask ourselves

why these thinkers ascribed such an urgent and pre-eminent significance precisely to the question of 'knowing'. What made it so important in the eyes of many of the post-Kantian idealists to demonstrate, for themselves and others, the untenability of skeptical objections to the possibility of knowledge? I have already suggested something of the answer to this question. And it takes us back directly to the controversies that surrounded the original reception of Kantian philosophy. Since some aspects of this reception certainly form part of the background of the post-Kantian discussion of the problem of knowledge, but did not substantively affect the actual development of the discussion, we can simply present the most important elements of this reception in its general outlines without going into particular detail.² In the first instance, I am more interested in sketching the key themes that feature in this period in the history of philosophy than in furnishing a full account of the period itself.

When we consider the full spectrum of initial reactions to Kant's philosophy as a whole, it is striking to note that one of the principal reservations concerning his theoretical philosophy was the suspicion that the ultimate implications of the latter only confirmed the skeptic's position with regard to knowledge. The origin of this suspicion is clearly closely connected with the perceptive and highly influential objections which F. H. Jacobi explicitly raised against the plausibility of Kant's philosophical doctrines. But for Jacobi himself, the problem of skepticism to which Kant theory was allegedly exposed was merely one amongst a whole range of difficulties which Kant's philosophical enterprise had incurred in Jacobi's eyes. It was simply one of the failings which essentially afflicted Kantian philosophy and explained why Kant's philosophical views, for Jacobi, ended up by presenting such a remarkable, and ultimately outrageous, contrast to the everyday conception of the world and the self which is entertained by all right-thinking human beings.³ Nonetheless, the critical objections which Jacobi raised against Kant's theory carried special weight for those who hoped to mobilize Kant's epistemic programme itself against the suspicion that it actually offered no convincing answer to skeptical objections concerning the possibility of knowledge. And Jacobi's objections carried the weight they did precisely because they could easily be interpreted as indications of a latent commitment to skeptical theses already implicitly contained in Kant's theory of knowledge itself.

Jacobi's own attack on Kant, in so far as it proved significant for the discussion of skepticism, was principally based upon two quite independent convictions which have both become sufficiently well-known. The first of them exploited the seemingly unavoidable consequences of Kant's distinction between 'appearances' and 'things in themselves' in conjunction with his claim that we can only properly 'know' the former.

² For a brief but clear and informative account of the issues and perspectives which shaped the discussion of skepticism during the last decade of the eighteenth century, and of the writers, many of them still barely known, who participated in this discussion, see Vieweg 1999 (esp. §§1.3 and I.4).

³ For Jacobi's understanding of Kant's thought, something which can only briefly and broadly be characterized here, see Horstmann 2004, 75ff.

Here Jacobi points out the following problem: if we can indeed only know appearances, while these must be sharply distinguished from things in themselves precisely with regard to knowability, these two claims, taken together, imply that things in themselves are unknowable for us. But since things in themselves were supposed to stand for what things ultimately or really are, that is, for what things are if considered as undistorted by the subjective contributions of the cognitive process, Kant's denial that we can know things in themselves is tantamount to abandoning the view that we can know anything that is independent of ourselves. That is the drift of Jacobi's argument.⁴ The skeptical conclusion which follows from these reflections is obvious: if we can actually know nothing that is independent of ourselves, then we cannot, for example, convincingly argue for the acceptance of the reality of the external world either, and this implies that we can never know, on Kantian premises, whether there is a subject-independent external world or not. But this is exactly what the skeptic also claims. Kant's restriction of knowledge to the realm of appearances is therefore incapable of countering skeptical doubts, for example, concerning the possibility of knowledge with regard to the external world. In this sense, Jacobi claims, Kant's concept of knowledge cannot escape the verdict of throwing the door wide open to the skeptic.⁵

Jacobi's second critical consideration derives from his reservations concerning Kant's treatment of the notion of the 'unconditioned'. According to Kant, as Jacobi interprets him, we can only recognize something as real if we can explain or demonstrate it. But all demonstrative explanation presupposes an unavoidable recourse to antecedent grounds or conditions. In order to avoid a perpetual regress in the chain of grounds or conditions, and thus in order to regard a knowledge or cognitive claim about the reality of something as finally and completely grounded, we require an 'ultimate' condition, or an 'ultimate' ground. This ultimate point of reference must be regarded as unconditioned, as immediate or unmediated in character, in short, as something that is not itself further explicable. It plays the role of justifying our knowledge and cognitive claims, even though it cannot itself be justified. The existence of the 'unconditioned' in this sense is thus a necessary condition for the knowledge or cognition of any real state of affairs. Up to this point, Kant's general epistemological perspective is entirely sound, for Jacobi, if indeed we can ascribe this line of thought to Kant himself. But now Jacobi indicates what he takes to be Kant's 'original sin': namely that this necessary representation of the unconditioned may only be ascribed the status of a regulative idea. Thus Kant can only recognize this representation as one that is indeed subjectively necessary, but remains objectively empty.⁶ This characteristically Kantian thesis disturbs Jacobi for reasons that appear to have little directly to do with skeptical

⁴ This perspective is already implicit in Jacobi's 'Supplement' on transcendental idealism. See Jacobi 1787.

⁵ A little later Salomon Maimon also presented a similar argument, in a modified but even sharper form, when he explicitly attempted to situate Kant's philosophy in much the same territory as Hume's skepticism. For Maimon's position in this context, see Breazeale 2003, 115–40, and Franks 2003, 141–63.

⁶ The basic features of this argument can be found in Jacobi 1789, particularly in the various 'Supplements' appended to the main text.

suspicions. Yet it is not difficult to see how Jacobi's argument here can lead to such suspicions. For if the representation of the unconditioned is indeed a necessary condition for the knowledge of the reality of actual states of affairs, while the representation in question has no objective reference or significance, and thus designates nothing that is actually given, then it appears impossible, on strictly Kantian assumptions, to characterize anything whatsoever as real. Thus Kant, we might suspect with Jacobi, makes himself complicit with the skeptic here too.

These and other of Jacobi's objections with respect to important elements of Kantian philosophy did not immediately lead, as I have already emphasized, to a direct engagement with the supposedly skeptical implications of the Critical philosophy, but they were nonetheless originally responsible for arousing a certain unease concerning the defensibility, or at least the clarity, of some fundamental assumptions of Kant's theory. This unease in turn prompted the emergence of new approaches the declared aim of which was precisely to set Kant's systematic work upon different and allegedly more viable foundations. Thus it was that thinkers now began the attempt to save the alleged 'spirit' of Kantian philosophy from its explicit 'letter', to proclaim the 'results' of this philosophy and to uncover the missing 'premises' still required to complete it. In this connection it is important above all to recognize the role of K. L. Reinhold, whose *Attempt at a New Theory of the Human Faculty of Representation* (*Versuch einer neuen Theorie des menschlichen Vorstellungsvermögens*) of 1789 initiated an entire series of similar attempts to provide a better foundation for Kantian philosophy.

It is also interesting to note that the principal discussion concerning the apparently skeptical implications of Kant's teaching was not generally conducted in terms of an explicit engagement with Kant's own theses, but rather sprang, in large part, from the critical controversies which were primarily connected with such attempts to provide an improved grounding for Kant's philosophy. And one work which was mainly concerned with criticizing Reinhold's attempt in this direction must be regarded as a key text in the discussion of skepticism. The work in question was an essay of 1792 by G. E. Schulze titled *Aenesidemus, Or Concerning the Foundations of the Philosophy of the Elements Issued by Prof. Reinhold, together with a Defence of Skepticism against the Pretensions of the Critique of Reason* (*Aenesidemus oder über die Fundamente der von dem Herrn Prof. Reinhold gelieferten Elementarphilosophie. Nebst einer Vertheidigung des Skeptizismus gegen die Anmassungen der Vernunftkritik*). The basic drift of the text is evident from the title: the 'critique of reason', i.e. Kant's philosophy, can only pretend to answer to the skeptic, and Reinhold's 'Philosophy of the Elements', which had been so widely lauded for improving the presentation of Kant's teachings, is entirely helpless before the objections of the skeptic, here personified by the ancient thinker Aenesidemus. It is not my intention here to identify or describe the specific theses of Kant or Reinhold which provoked this critical diagnosis, or to decide whether Schulze's criticisms can actually be justified. It is merely important to recognize that many of his philosophical contemporaries did indeed take Schulze's arguments seriously and evinced considerable respect for his substantive contribution to the debate.

These contemporaries certainly included Fichte and Schelling, as we can see from their early writings, and a few years later, at the beginning of his philosophical career, Hegel also felt it was worth engaging in some detail with Schulze's skepticism.

If we bear this background in mind,⁷ it is easier to understand why the interest in questions concerned with grounding knowledge, with obviating skeptical objections, with elucidating the possibility of knowledge, appeared philosophically self-evident in the eyes of all those who felt inspired in one way or another by the results and the spirit of Kant's philosophy. For only if we can defend a convincing position with regard to these questions, can we also hope to ascribe a certain objective character to the implications and consequences of a philosophical reflection pursued in an expressly systematic form. The question of 'knowledge' was more than an object, and one amongst others, of merely intellectual curiosity since it concerned the foundations and the possibility of philosophy itself.

II

As yet the manner in which I have spoken of 'knowledge' and of the 'skeptical' challenges raised against its claims have not been fully or adequately clarified. First of all, therefore, I must attempt to specify both questions more precisely in order to help articulate the framework in which the discussion concerning knowledge and skepticism belongs if we are to relate it directly to the theoretical scene of post-Kantian idealist philosophy.

If someone attempts to persuade us that we have no good reason for assuming that we do indeed know something, such a person can be said to defend a skeptical position. Skeptical positions, of course, can be defended in relation to the most various issues and areas. The most philosophically significant skeptical positions in the history of modern philosophy have proved to be those discussed principally under such headings as 'the reality of the external world', 'the existence of the self', 'the freedom of the will', and 'the existence of God'.

What all skeptical reflections have in common is the fact that they argue on the basis of a specific concept of 'knowledge' and then proceed to exploit specific difficulties which arise, or allegedly arise, in connection with the concept in question. As far as the concept of knowledge is concerned, the skeptic generally assumes an interpretation of epistemic truth which effectively goes back to Plato. According to this interpretation, we can rightly claim to know that a state of affairs exists or is the case if we possess a justified true belief regarding the existence of this state of affairs. Thus, on this interpretation of 'knowledge', one knows that it is raining if (1) he believes or thinks that it is

⁷ A more complete picture of the post-Kantian discussion of skepticism and the substantive connections involved is furnished by the contributions of Daniel Breazeale, Thomas Grundmann, Marcelo Stamm, and Achim Engstler on the theme of 'German Skepticism after Kant', in van der Zande and Popkin 1998. G. E. Schulze's *Aenesidemus*, and Fichte's 'Review' of 1794, are translated in di Giovanni and Harris 2000.

raining, if (2) he has acceptable reasons for his belief that it is raining, and if (3) it is in fact raining.⁸

Now it appears quite evident to the skeptic that this very analysis of knowledge implies that we cannot actually know many things that we usually claim to know perfectly well. And above all, so the defender of the skeptical perspective will claim, we cannot uphold, on this analysis, claims that suggest we know something about the existence or intrinsic character of a reality that is independent of the subject. The skeptic argues that we cannot know any such thing because it can never be established in such cases that the relevant conditions have been actually fulfilled. In support of this view, the skeptic generally adduces a series of observations designed to show that we can never exclude the possibility that we may be deceived about anything and everything. This reservation concerning the possibility of deception is supposed to apply to immediate (direct) or non-inferential knowledge and to mediated or inferential knowledge alike. The possibility of deception is said to hold for non-inferential or immediate knowledge precisely because—and we hardly needed to wait for Descartes to appreciate this—the human senses can indeed deceive us. And it holds for inferential knowledge because every premise to which we appeal to justify some specific knowledge claim is itself either an immediate (non-inferential) case or a mediated (inferential) case of supposed knowledge. If we claim that the premise is known non-inferentially, it is exposed to the possibility of deception to which all immediate knowledge is allegedly subjected; if it is known inferentially, then it must depend upon a further premise which, in turn, could only be known either inferentially or non-inferentially. Since this alternative holds for every possible further premise, we must either accept an infinite regress in the justification of our knowledge claims, or we must claim that some premise or other is indeed known non-inferentially. Whichever alternative one chooses here, we must abandon the claim that there is any ‘knowledge’ which is absolutely free from the suspicion of deception. In this situation, the skeptic concludes that there is no knowledge, regarding either the existence or the character of an external world independent of the subject, which is immune to skeptical doubt. On the contrary, all our knowledge claims in this regard rest on ultimately indemonstrable presuppositions that we simply accept or believe.

This skeptical diagnosis has, of course, repeatedly provoked a host of responses which have led in turn to several different theoretical accounts of knowledge. In the last analysis, however, all of these responses can be reduced to two fundamental approaches. The first typically defends the conviction that the possibility of knowledge is in fact closely connected with the problem of justification through reasons, while the second typically abandons that conviction, and thus assumes that we do not need to address the problem of justification in order to secure the possibility of knowledge.

⁸ Whether or not these three conditions are sufficient for characterizing something as knowledge is a controversial question (see the discussion surrounding the so-called ‘Gettier paradox’), but it has rarely been contested that they are necessary conditions of knowledge.

The theories which are articulated within the framework of the first approach may be described for obvious reasons as 'justification-oriented' ones, while the theories which are indebted to the second approach may be described, for reasons that will become clearer later, as 'grounding-oriented' ones.

With regard to theories of knowledge which are oriented to the central role of the notion of justification, we can in turn distinguish two groups, the first of which can be described as 'foundationalist', and the second as 'relativistic' or 'contextualist' in character.⁹ If we adopt this taxonomy, foundationalist theories are those which claim to demonstrate, by one means or another, that there are examples of non-inferential knowledge that are immune to any possible deception (in the case of Descartes: the examples of knowledge of ourselves and knowledge of God). Relativistic or contextualist theories, on the other hand, lower or moderate the standards of justification that are allegedly required for knowledge. According to this perspective too, we must indeed be capable of justifying any knowledge claim by appeal to reasons that exclude the possibility of deception. But, at the same time, it denies that this condition of justification is only fulfilled if the relevant reasons can be traced back to some non-inferential knowledge that is immune to doubt and can thus exclude the possibility of deception. On this perspective, rather, the condition of justification is already held to be fulfilled if our reasons merely exclude those possibilities of deception which are relevant to the specific context in which the knowledge claim is made.¹⁰

In contrast to such justification-oriented theories, those associated with the other basic approach attempt to show that there are ultimate and irreducible factors which can explain the possibility of both inferential and non-inferential knowledge once we show how these factors are fundamentally constitutive for knowledge. The expression 'constitutive for knowledge' here is not intended to suggest some specific process or other which generates or gives rise to some substantive knowledge of something or other. On the contrary, it is supposed to designate the sum of those conditions which must be fulfilled if we are to speak of knowledge at all. Examples of just such epistemically constitutive factors would include the 'I' that posits its own being in Fichte, or the 'spirit' which actually recognizes itself as such in Hegel. What these grounding-oriented approaches have in common is the fact that they clearly do not question the possibility of knowledge. But they do not undertake to challenge the conclusions of the skeptical argument by attempting to refute the premises of the latter. These two approaches differ principally from one another, first, with regard to the way in which they derive and present the relevant constitutive factors of knowledge in each case (Fichte's concept of 'reflective abstraction' and his notion of genetic explanation, and Hegel's conception of

⁹ This rather rough and ready distinction does not pretend to capture the highly nuanced and differentiated classifications that have now become current in contemporary discussions of the philosophy of science and knowledge. In relation to this discussion see e.g. Kompa 2001.

¹⁰ Contextualist theories of knowledge are currently debated principally within the analytically oriented Anglo-American field of philosophy, and are associated with the names, amongst others, of Alvin Goldman, Fred Dretske, David Lewis, and Keith DeRose.

'self-accomplishing' skepticism) and, secondly, with regard to the way in which they address the question concerning the epistemic justification of these factors (Fichte's 'intellectual intuition', and Hegel's conception of speculative 'self-knowledge').

It is thus easy to see that justification-oriented theories and grounding-oriented theories of knowledge react to the skeptical argument from infinite regress in very different ways. Both types of justification-oriented theory hope to refute the skeptic directly by attempting to terminate the regress diagnosed by the skeptic, and to do so by identifying some epistemically privileged factor or feature in our knowledge. For the foundationalists these factors are furnished by cases of non-inferential knowledge immune to deception of any kind, while for the relativists they are defined by the concept of relevant context. The two approaches differ from one another insofar as the foundationalist version attempts a positive demonstration of the existence of non-inferential knowledge, whereas the relativistic programme is primarily interested in firmly securing the requirements which conditions of justification must satisfy. The proponents of the grounding-oriented approach do not regard these strategies as particularly promising or persuasive. That is why they avoid confronting the skeptical argument directly, and attempt instead to undercut the argument in question. They do so by conceiving knowledge in such a way that its sense and significance can already be seen to imply the existence—and that is to say, the reality—of certain fundamental features and factors (the I, the world, life). In this way they attempt to uncouple the question how or whether knowledge is possible from those problems concerning its immunity to doubt or its exposure to infinite regress which the skeptic had raised in the first place.

I wish to defend the thesis, which will be supported by further observations in what follows, that the most important of the post-Kantian idealist approaches to the problem of accounting for knowledge should be seen as attempts to avoid the skeptical challenge precisely by developing what I have called 'grounding-oriented' positions.¹¹ This can be shown relatively clearly above all with respect to the works of Fichte (and, in his wake, of Schelling), but also I think in the case of Hegel. Although none of these three thinkers ever seriously doubted the capacity of the grounding-oriented approach for neutralizing the skeptical challenge, in some respects they differed considerably from one another in their respective ways of executing and fulfilling the general philosophical programme. Whereas the problem of skepticism only really interested Hegel and Schelling at certain periods of their development, it continued to preoccupy Fichte for much longer. This is already evident from the numerous versions of the so-called *Wissenschaftslehre* which Fichte created for this purpose, all of which are concerned, in a certain sense, with clarifying and establishing the possibility of knowledge as such.

¹¹ In fact this thesis is intended to refer only to the post-Kantian idealist positions, without thereby claiming that there is a fundamental difference between these approaches and the position of Kant. How Kant's theory of objectively valid knowledge should be situated in the conflicting field of 'justification-oriented' and 'grounding-oriented' anti-skeptical strategies is an extremely interesting and important question which I cannot attempt to address here.

The various versions of the *Wissenschaftslehre*, according to the prevailing view, can roughly be divided into two groups. The first group comprises the versions which Fichte composed before 1800. These versions, as I hope to show, are all characterized by the fact that Fichte, despite his best intentions, does not yet succeed in emancipating himself from a foundation-oriented approach. The second group consists of the versions Fichte developed after 1800, and in these he attempts to make a radical new beginning and now defends, by a variety of means, a consistent grounding-oriented approach.

III

As is well known, Fichte's *Foundations of the Entire Wissenschaftslehre* (*Grundlage der gesamten Wissenschaftslehre*) of 1794/5 is merely the first of many attempts to solve the contemporary problem of establishing a theory of knowledge that is capable of withstanding all skeptical objections and reservations.¹² This initial approach is influenced, on the one hand, by Fichte's endorsement of Reinhold's position in the argument between Schulze and Reinhold concerning the possibility of securing a philosophical foundation immune to skepticism through recourse to a first or fundamental principle (*Grundsatz*) of philosophy, and, on the other, by his acceptance of Jacobi's claim that there is basically no sense in trying to refute the arguments of the skeptics by pointing out mistakes in the arguments which they employ.¹³ Fichte's support for Reinhold, and thus for the necessity of a first principle 'of all human knowledge', is grounded in the double assumption (1) that we can only properly be said to know assertions (judgements or propositions)¹⁴ which exhibit the character of certainty (*Gewissheit*), or those which thus express what is actually and certainly the case; and (2) that a regress in the process of justification with regard to such assertions can only be avoided if we are

¹² Since this first presentation of the *Wissenschaftslehre* is the only more or less developed version of his general philosophical project which Fichte published in his own lifetime, it is hardly surprising that the 200-year-long debate concerning the merits and deficiencies of Fichte's position has concentrated principally on this presentation of his thought. The consequence has been a growing, and now rather unmanageable, body of literature concerned both with the overall project itself and the most various aspects of this early *Wissenschaftslehre*, and which for practical reasons of space will not be mentioned in any detail here. But the contributions of Fichte's contemporaries, such as the early Schelling, the Hegel of the Jena period, and Jacobi, to the critical examination of Fichte's early theory nonetheless remain extremely instructive. For information on the more recent relevant literature, see the bibliographies in Zöllner 1998 and Klotz 2002. The interpretation of Fichte's argument in the opening section of the first version of the *Wissenschaftslehre*, as presented in this section of my chapter, represents a slightly modified version of the outline in Horstmann 2000, 121–6.

¹³ For an illuminating discussion of the controversy between Reinhold and Schulze, and its importance for the early Fichte, see Breazeale 1980–1, 1982, as well as Lauth 1989. For Jacobi, see the texts already mentioned.

¹⁴ In the following discussion I do not draw a distinction between the terms 'judgement', 'assertion', or 'proposition'. Fichte normally uses the word 'Satz' to designate what today we would call a judgement or proposition according to context. According to their grammatical form, Fichte's 'propositions' are supposed to represent subject-predicate propositions, or ones which can be reduced to subject-predicate propositions.

able to furnish some principle or fundamental proposition which is 'utterly unconditioned' (*schlechthin unbedingt*) (as Fichte puts it), one, that is, which cannot be derived from any other principle, and which, for its part, is of such a kind that it alone guarantees the utter certainty, and thus the indubitability, that is, the immunity to skeptical objections, of a given proposition. This conviction, which Fichte shares with Jacobi, leads him to the claim that the unquestionable certainty of a proposition can never be demonstrated discursively (by appeal to purely conceptual considerations) or intuitively (by appeal to any sensuous perception), and that, on the contrary, the ground of unconditional certainty can only be found in the constitution of self-consciousness itself.

In fact Fichte never really made it clear why the first principle that he was seeking was supposed to have anything to do with the specific features of self-consciousness. It seems plausible to assume that Fichte was impressed by either one or both of the following considerations: (1) the first principle is supposed to represent the principle of all knowledge. Knowledge is the result of a conscious activity and to that extent represents a state or condition of consciousness. States of consciousness are states of a subject that is conscious of the latter, and must therefore possess a relevant I-consciousness or self-consciousness. Such a subject is precisely what we describe as an 'I' or a self. Thus we must already, for reasons which are connected with the concept of knowledge, be able to set knowledge in relation to self-consciousness. We seem justified in ascribing such considerations to Fichte by the way in which he actually introduces the 'I' in the so-called *First Introduction to the Wissenschaftslehre*. And (2), one of Kant's essential insights, which effectively belongs in the Cartesian tradition, is that all knowledge claims are ultimately grounded in the activity which is made possible by what he calls the 'transcendental unity of apperception' or the 'transcendental I'. Now we cannot contest this insight, which is indeed entirely convincing. But Kant himself never succeeded in rendering the relationship between knowledge and the 'I' sufficiently transparent. If we wish to preserve Kant's insight here, it must be transformed into the demonstration that the first principle of all knowledge itself is nothing but a structural feature of self-consciousness as such. And this reflection, which is grounded in the actual historical context, would certainly reflect the Kantian roots of Fichte's thought which he himself would repeatedly emphasize. Thus for Fichte—in the context of this, the first formulation of the *Wissenschaftslehre*—the ultimate intention of a theory of knowledge capable of withstanding all skeptical objections can only be realized if we can discover a first principle that is grounded in self-consciousness itself.

Fichte tells us more precisely what this first principle must look like in the opening section of his *Wissenschaftslehre*. Amongst the many problems connected with the discovery of this fundamental principle, the first one is to identify an appropriate method by means of which the principle in question can be derived. And here Fichte proposes a procedure which he calls 'abstractive reflection'. This kind of reflection, according to Fichte, takes its point of departure from a proposition which everyone regards as unquestionably certain, from a so-called 'fact of empirical consciousness'. From this

fact, this indubitable proposition, the process of reflection isolates (abstracts) the elements which belong to the content of such a proposition, i.e. belong to that about which the proposition asserts something. What is supposedly left, after this abstraction, is simply the form of the proposition which consists precisely in affirming the ascription, or non-ascription, of a predicate to a subject. Now Fichte seems to believe that, with respect to indubitably certain propositions, the relationship between subject and predicate is a necessary one. And reflection upon (and here that means: the explanation of) the necessity of this connection is supposed, according to Fichte, to lead to the self-positing activity of the I, or what Fichte calls the 'Act' (*Tathandlung*), which functions to that extent as the necessary condition of the certainty of a proposition, and is thus properly qualified as the first principle of knowledge.¹⁵

The proposition which Fichte selects as the point of departure for his 'abstractive reflection' is the logical law of identity in the form ' $A = A$ '. In a rather simplified fashion, we can summarize Fichte's general line of argument as follows: (1) the proposition ' $A = A$ ' is rightly regarded as utterly certain, i.e. certain without recourse to any further grounds and thus as intrinsically certain. This fact alone already shows that we have the capacity (the faculty) of claiming something as certain without reference to any further grounds, or, in Fichte's own terminology, of 'positing' something absolutely (*etwas schlechthin zu setzen*). And (2), what is affirmed with the proposition ' $A = A$ ' is, according to Fichte, a conditional or hypothetical judgement that claims: 'if A exists, then A exists'. With this move in his interpretative chess game, which transforms a categorical judgement concerning the identity of A with itself into a hypothetical judgement concerning the being (the existence, the reality) of A, Fichte can claim as a result that the proposition ' $A = A$ ' actually determines something about the existence of A, namely that A is under the condition that the being of A is asserted (or posited); (3) Fichte initially analyses the kind of connection between the 'if-part' and the 'then-part' of a hypothetical proposition. He concludes that the connection between the two parts must be a necessary one if it is correct that the whole proposition is utterly certain.¹⁶ Necessary connection is thus constitutive for utter certainty. Fichte describes this necessary connection as 'X' and, since it represents for him the necessary condition for the utter certainty of a proposition, concentrates his efforts on the explanation of its

¹⁵ We should note two points in this connection: (1) the fact that Fichte is principally concerned about questions regarding the certainty of a proposition—and not e.g. with truth or verifiability—already indicates that he is primarily interested in overcoming skeptical objections against the very possibility of knowledge. (2) The procedure of 'abstractive reflection', as explicated in the *Wissenschaftslehre* of 1794/95, no longer plays any special role in Fichte's later writings. This may derive from the fact that, even in the first version of the *Wissenschaftslehre*, it is extremely difficult to reconcile Fichte's own description of this procedure with the way in which he actually arrives at his first principle.

¹⁶ Fichte appears to assume that the utter certainty of a proposition implies that the predicate of the proposition necessarily applies (or fails to apply) to the subject. He presumably endorses this view because, in cases where the predicate applies only contingently to the propositional subject, we can no longer claim absolute certainty, i.e. a certainty that cannot be grounded in anything else. For in order to regard a contingent proposition as certain, we should already require a ground.

possibility. (4) Fichte begins with a far-reaching claim, which he never actually clarifies or explains more precisely, concerning the allegedly necessary implications which also follow from his interpretation of the principle of identity. For he declares that everything that is constitutive for the utter certainty of a proposition is also posited absolutely. Since, according to (1), the proposition 'A is A' is posited as utterly certain, and since, according to (3), 'X' is constitutive for the utter certainty of 'A is A', then 'X' too, Fichte claims, 'is posited absolutely, and without any ground whatsoever' (GA i/2. 257). (5) Since 'X', according to (3), expresses a necessary connection, or relation, between A (the propositional subject) and A (the propositional predicate), then it follows, in accordance with (4), as applied to A, that A is also necessarily posited. For, as Fichte explains, if a specific relation is posited, then the relata which are meant to be related by the relation must be posited too. (6) The proposition 'A is A' affirms the identity of A with itself. It must hold, therefore, that A—*on condition that it is posited*—is necessarily posited. On this conditional reading of Fichte's argument, the proposition 'A is A' signifies the following: if the A (the propositional subject) is posited, then this A (the propositional predicate) is (exists). We should note here that Fichte's inference from positing to being certainly has its controversial aspects, although these cannot be pursued here. (7) 'X', and therefore also A, are acts of positing on the part of the subject that affirms (posits) the proposition 'A is A': or, in Fichte's terminology, they are 'posited in the I, and through the I' (GA i/2. 257). (8) The utter certainty of the proposition of identity is grounded, therefore, in the positing activity of the I (which in this case posits identity), an activity which consists precisely in postulating the being of what has been posited as identical.

All of Fichte's reflections up to this point¹⁷ are obviously supposed to show the following: if any proposition is utterly certain, it is so only under the condition that we also affirm (posit) its content as existing. And this by virtue of the circumstance that being belongs to the presuppositions of certainty that are posited by the I itself. It is already evident, by this point, that Fichte intends to make the being-positing power of the I the basis of a conception of knowing which is immune to skeptical objections. The fact that no skeptic has ever been particularly impressed or forcibly convinced by the anti-skeptical potential of this argument is due to two things: (a) Fichte's reflection, even if we accept that it is not internally incoherent, can only demonstrate, in the best case, that being (reality) is a condition of certainty, but it cannot show that something actually exists (is real); and (b) it introduces being (reality) as itself connected to a condition, namely to the being-positing activity of the I, but it cannot suffice to establish the being, the reality, of the I itself.

Fichte seems to have been fully aware of the limited anti-skeptical force of the preceding reflections and considerations. In any case, he himself explains that the result that has been obtained from steps (1)–(8) is not yet sufficient to demonstrate

¹⁷ Fichte's argument is presented in 1–5 of the first section of the *Foundations of the Entire Wissenschaftslehre*.

the first unconditioned and fundamental principle of all knowledge. And that is because, in the preceding course of these reflections, we have arrived at the I as the guarantee of the absolute certainty of a proposition only on the basis of an empirical fact, namely the proposition 'A is A' which has been presupposed as utterly certain. Fichte now rightly observes that the 'I am', as the condition of the certainty of an empirical fact, itself merely possesses the status of an empirical fact (*Tatsache*)—if the utterly certain proposition were not given, one would never be able to affirm the utter certainty of the 'I am'. Up to this point the 'I am' has 'only been grounded on a fact, and possesses no other validity than that attaching to a fact'. But in order to discover the first utterly unconditioned principle of all knowledge, we must establish not only the utter certainty, but also the unconditional certainty of the 'I am', one, that is, which does not depend upon the presence of a fact at all. In other words: we must be able to answer the question how the utterly unconditioned certainty of the proposition 'I am' is possible.

It is Fichte's reflections on this question¹⁸ which lead to his conception of the I as an 'Act' (*Tathandlung*). If we try to summarize his answer to this question, we express his argument roughly as follows. From the line of thought we have already presented (steps (1) and (7)) we know that the I has the capacity to posit something absolutely in the I. But in order to be able to posit something absolutely, the I itself must be posited. We have also already seen that the absolute positing of the I consists in the activity of positing being. If this is so, and if the 'I am' is to depend on nothing else as its condition, it seems plausible to interpret the I as the product of its own positing activity, since it would otherwise be quite impossible to explain its being. Now this, in turn, is supposed to imply that we must think the I as an activity which posits its own existence insofar as it is active. Fichte expresses this in section 6 of the first division as follows: the I 'is at the same time the actor and the product of the act; the actor, and that which the activity brings forth; act and deed are one and the same; and the "I am" is therefore the expression of an Act' (i/6. 259). For Fichte, an Act is not supposed to be a fact, that is, something which is simply already discovered as given, since the Act is logically and ontologically prior to any facts insofar as it ultimately constitutes (posits) everything which can be a fact for an I. The I, understood as Act, is supposed to be something absolutely posited precisely because it posits itself, and this self-positing constitutes its essence and guarantees its being, its reality. Again, as Fichte says: '*That whose being (essence) merely consists in positing itself as being, is the I, as absolute subject*' (GA i/7).¹⁹ This means, for Fichte, that the I, so understood, displays all the characteristics which make it an appropriate candidate for the first utterly unconditioned principle of all knowledge. Fichte tries out various formulations for expressing this first principle in a really adequate fashion. His most accessible formulation is certainly

¹⁸ They are presented in 6–10 of the first section of the *Wissenschaftslehre*.

¹⁹ Fichte claims that it is this formulation which makes it 'completely clear in what sense we are using the word "I" here'.

the one he furnishes at the end of section 10: '*The I originally posits its own being absolutely*' (i/10. 261).²⁰

This brief summary of what Fichte himself regarded merely as the basic outline of a plausible argument for establishing his first fundamental principle does not pretend to have reproduced his reflections in any adequate detail. But hopefully it will serve to indicate the general strategy which he adopts in pursuit of his goal, namely to lead us towards the insight that the I must be conceived as self-positing activity. For Fichte, this insight is not merely significant in its own right, but also provides both the basis for the epistemic justification of a logical law, namely the principle of identity, and the foundation for deriving the category of reality.

For reasons which are either untenable,²¹ or at any rate are exceedingly obscure, Fichte then introduces two further principles of all knowledge, which are not, like the first fundamental principle I have discussed, utterly unconditioned, but are supposed in each case to be unconditioned in one specific respect. Fichte attempts to justify the introduction of two further principles on systematic grounds, although these principles can only be described as unconditional in a qualified respect, by exploiting his own distinction between the form and the content of a proposition.²² According to Fichte, every proposition (judgement) can be treated as either conditioned or unconditioned in relation to its content, or to its form, or to both. If a proposition is unconditioned in either or both of these respects, then it can be described, in Fichte's terminology, as a fundamental principle (*Grundsatz*).²³ These considerations imply, in an entirely trivial sense, that we must bring three such principles into play here. With his second and his third principle Fichte effectively attempts to neutralize two problems which might threaten to undermine the anti-skeptical thrust of his own theory of knowledge. The first problem can be expressed in terms of the following question: even if we assume that Fichte's first principle is actually valid, how can we still affirm that the world to which our knowledge claims are directed can properly be regarded as an ensemble

²⁰ The methodical path by which Fichte introduces the I as the original Act, which I have just sketched, is far from immediately, or even indirectly, convincing. Indeed Fichte soon criticized it himself, as we can clearly see from his correspondence. He appears to have been particularly dissatisfied with this attempt to connect the conception of the I as Act with the question concerning the first principles of knowledge.

²¹ It would be untenable e.g. for Fichte to introduce these two further principles simply because his first principle is so impoverished that it is unable to establish the necessary relation to an object that is required by any knowledge claim that furnishes more than tautologies. In this case, the further principles could not claim anything more than the status of ad hoc assumptions.

²² It is by no means entirely clear how the distinction between the form and the content of a proposition is to be understood precisely, especially in the case of the fundamental principles. Fichte's own attempt to explain how this distinction should be understood can be found at GA i/2. 121 and 257.

²³ The generation of these fundamental principles leaves a large number of unresolved questions which Fichte never really attempted to answer. He did address the most pressing of these questions, namely why there can only be a single absolutely unconditioned principle, in his short piece *Concerning the Concept of the Wissenschaftslehre* of 1794 (GA i/2. 114ff.; English translation in Fichte 1988). Yet the grounding he offers there cannot simply be applied to other significant questions, such as why e.g. there cannot be more than one principle which corresponds to the criteria for a principle that is unconditional either with respect to form or to content.

of internally related objects that is independent of ourselves? Or to put the question in another way: how can we retain the concept of an object that is distinct from the knowing subject if, in accordance with our first principle, all being (reality) is a product of the self-positing I? The second problem concerns the possibility of an epistemic relationship between subject and object on Fichtean premises: how must the character of the I and its object be conceived if they are to stand in a relation of knowledge to one another?

Fichte responds to these problems, as we have observed, by introducing two new principles which are presented as unconditional acts of positing on the part of the I. According to Fichte, the I possesses, in addition to the capacity for self-positing that is captured in the first principle, the further capacity of positing a non-I freely and simply (without any further ground). That is to say, it has the ability, through what Fichte calls an 'absolute act' (*absolute Handlung*), to posit something as non-I that is opposed to the I. Fichte's second principle codifies this act of 'counter-positing'. Finally, the I is further characterized by a third capacity, that of freely and simply positing the divisibility of the I and the non-I. The third principle specifically captures this notion of divisibility.²⁴

It is not difficult to grasp what Fichte is attempting to accomplish with the introduction of his second principle. It is supposed to do justice to our inextinguishable everyday conviction that there is an external world independent of ourselves, the objects of which are outside of and distinguished from us, and to which we can relate in terms of knowledge and action alike. But for Fichte this conviction is justified not because an external world independent of ourselves compels us to understand it as characterized in such and such a way. On the contrary, it is justified because it belongs to the distinctive structure of the I to organize its world dichotomously through the subject-object distinction or the opposition of the I and the non-I. It is more difficult, on the other hand, to grasp the significance which Fichte wishes to ascribe to the third principle of divisibility. The motivation for introducing it is obviously to present the non-I not only as the negation of the reality which the I claims to posit in positing its own being, but rather to ascribe independent reality, independent being, to the non-I. In pursuing this purpose by recourse to the concept of divisibility, Fichte appears to make the implicit but rather strange presupposition that being or reality should be regarded as a kind of

²⁴ The categories of negation and determination (Kant's category of limitation) are supposed to follow from the second and third principle in each case. Fichte claims that these two principles are formally subordinate to the first principle because they are only unconditional in part. This claim would appear to rest on the assumption that, with regard to the second principle, there is no direct conceptual relation between the acts of positing and opposing, but only one that concerns its act character. In the case of the third principle, however, there is a conceptual reason for introducing the concept of divisibility, but none which lies in the act character of the two acts. On the basis of this assumption Fichte characterizes his second principle as conditioned with regard to content, and the third principle as conditioned with regard to form. This rather artificial explanation of the respective status of the principles clearly indicates how difficult Fichte found it to accommodate the fundamental steps of his anti-skeptical programme within the framework of the 'philosophy of principles' which originated with Reinhold.

quantity, as something given in degrees (intensively considered) or parts (extensively considered). Once we accept this presupposition, then we cannot in fact avoid bringing something like Fichte's third principle into play. For the assumption that we have to conceive reality as a distributable plurality, together with the notion that there are real objects possessed of an existence independent of the subject, means that it is necessary, within the Fichtean model of positing, to identify a factor responsible for distributing reality between the I, understood as the knowing subject (and not as absolute self-positing ego), and the non-I, understood as the object that it is to be known.²⁵

However matters stand with regard to the precise details, the Fichte of the *Foundations of the Entire Wissenschaftslehre*, at least, clearly believes that the three principles which he proposes have enabled him to accomplish his twofold objective: (1) to provide a non-dualistic explanation of the most fundamental acts and concepts that are constitutive for reality, and (2) to develop an anti-skeptical argument by conceiving the I as an activity which posits its own being, and thereby establishing the relation to reality that is indispensable to his concept of knowledge (for in the final analysis it is supposed to hold that whatever we know must be the case).

IV

The fact that Fichte abandoned this initial approach of 1794/5 fairly soon afterwards is clearly documented both by his subsequent published writings and, especially, by the transcripts of his lectures from the period between 1796 and 1799. The reasons for the various modifications which he felt compelled to make to his original approach are not particularly difficult to identify since Fichte himself occasionally refers to them more or less explicitly. All of them are connected in one way or another with his emerging reservations concerning the general procedure for establishing his first two basic principles which he had adopted in the version of 1794/5. This procedure must be interpreted in terms of the strategic goal he originally pursued. In his initial version the strategic plan was to begin from a fact, a fact of consciousness, and then, through a process of 'reflective abstractive', to uncover the conditions which made this fact possible, and to interpret the entirety of these conditions in terms of a conceptual construct which is to be identified with the Act that is the 'I'. The Fichtean conviction underlying this strategic reflection thus surely consisted in the insight that the best way of responding to skeptical objections against the very possibility of knowledge was to demonstrate an intimate and intrinsic connection between knowledge and self-consciousness and to show that it is precisely the existence of this connection which is necessary in any case for the possibility of knowledge, and is probably also sufficient to establish this possibility. If this diagnosis is correct, then it is already quite evident that Fichte is not

²⁵ For the details of Fichte's argument with respect to the second and the third principle, see Horstmann 2004, 84ff.

primarily interested in the question concerning how we can secure the reality of the varied content of knowledge as something independent of the subject. It seems that he is principally interested in the question as to what conditions must be fulfilled to secure the possibility of the presence of the propositional attitude involved in 'knowing'. And for Fichte the subject-independent reality of what is known belongs precisely amongst these conditions. This implies, to put it rather pointedly, that even in this early period of his thought Fichte is not particularly interested in the skeptic who grounds his skeptical position on our capacity to cast doubt upon the content of knowledge, or who plagues himself and others with the same perennial anxiety that whatever we may claim to be the case might indeed just as well not be the case. The skeptic with whom Fichte feels he must engage, on the other hand, is obviously the one who claims that the (mental) state, attitude, or posture that we call 'knowing' is a simply unrealizable chimera because precisely the relationship to reality, that which ensures that something is indeed 'the case', is never guaranteed. As far as Fichte's procedure is concerned, this means that the possibility of appealing to such a fact of consciousness constitutes an essential element of his entire strategy. In view of the anti-skeptical intentions of Fichte's basic project, it is obvious that this fact itself must be a case of knowledge, that is, must be something that is immediately, or unconditionally, certain.

Fichte's demonstration of the alleged connection between knowledge and self-consciousness—a demonstration largely independent of his general procedure—is already burdened, at least in the form in which Fichte presents it, with considerable difficulties with regard to the underlying analysis and argument, difficulties which a skeptic could also have attempted to exploit in favour of his own cause. But even if we disregard this, the real problem that besets an anti-skeptical strategy of this kind lies elsewhere. The real problem is that something like the reality of knowing or, perhaps better put, that the actual possession or presence of knowing is here presupposed *as a fact*. For this is precisely what Fichte means to say when he speaks of unconditionally certain facts of consciousness: we can proceed from the assumption that there are indeed cases of knowledge, and it is only because of this, only because this is itself a fact, that we must attempt to explicate the latter in adequate terms. But it is precisely the presupposition of the reality of knowing that turns the Fichtean programme, as it is presented in 1794/5, into an extremely blunt weapon for battling the skeptic in the cause of 'knowledge'.²⁶

For as far as the skeptic is concerned, nothing is more obvious than the temptation to challenge this very presupposition, either to identify a vicious circle in the Fichtean argument, or to claim that Fichte's approach argues past and thus misses the point of

²⁶ This certainly does not imply that this Fichtean presupposition and the programme of the first version of the *Wissenschaftslehre* that operates with it is without any cognitive value whatsoever. The conditions under which knowing is possible are naturally of central importance for any philosophical theory of knowledge. It is simply meant to indicate that Fichte's early approach cannot be regarded as particularly effective in the context of general anti-skeptical strategies.

the full skeptical challenge. The skeptic could frame the charge of circular reasoning as follows: if I doubt the possibility of knowing, this implies that there is not a single case where we can indubitably demonstrate the fulfilment of those conditions which are required to establish real knowledge. To doubt the possibility of knowing is just to assert the unfulfillability of the conditions required for knowledge. The skeptical doubt, in other words, does not even have to be directed against the conditions themselves, but need only be mobilized against the idea that they can ever be fulfilled. If you wish to refute my position, so the skeptic further argues, it is not enough therefore to presuppose (cases of) knowledge as actually present and then proceed to reconstruct the conditions under which such knowing is possible. For, speaking as a skeptic, I have no problem about inferring from reality to possibility. One would thus also have to show, in addition, that these conditions are actually fulfilled. As a skeptic, therefore, I have problems with the following argument: (1) There is actually (some case or cases of) knowledge. (2) Knowledge is therefore possible. (3) In order to explain the possibility of knowledge, we must refer to such and such conditions. (4) If these conditions are fulfilled, then there is knowledge. (5) There is indeed knowledge (as in (1) above), and the relevant conditions are thus fulfilled. Apart from the fact that the argument is already problematic insofar as (4) draws conclusions about reality from the conditions of possibility, it also fails as a genuinely anti-skeptical argument since it presupposes precisely what was supposed to be shown: namely the reality of knowledge. To put the matter in another way: even as a skeptic, I do not deny the validity of the conditional proposition: if knowledge is to be possible, such and such conditions must be fulfilled. But I definitely deny that this conditional entitles us to claim I can know that such and such conditions are fulfilled, and knowledge therefore (really) exists. Such a claim—which is already formally invalid since it can only be derived from the conditional by ignoring the logical principle of contraposition—is untenable, above all, since we cannot demonstrate that these conditions are actually fulfilled without already presupposing actual knowledge (here, either that these conditions are fulfilled, or that a proposition is certain because it fulfils these conditions). The claim is not properly justified by anything in Fichte's argumentation and reveals the circularity of his attempt to prove, against the skeptic, that there is such a thing as knowledge. It may therefore well be true that knowledge is impossible without recourse to something like Fichte's conception of the 'I' or the original 'Act', but it does not follow, with this conception or on the basis of this conception, that there actually is such a thing as knowledge.

But apart from raising the suspicion of circularity, the skeptic could also object that Fichte's strategy in the 1794/5 version of the *Wissenschaftslehre* effectively avoids rather than confronts the real skeptical claim. For even if we concede that Fichte's philosophical approach makes some contribution to answering the question *how* knowledge is possible, this approach yields no significant response whatsoever to the quite different and authentically skeptical question *whether* knowledge is possible. The sort of genetic explanation of knowledge which Fichte undertakes to provide is thus beside the point if we are attempting to decide upon the justification of skeptical claims in the first

place. After all, one would hardly decide on a positive answer to the question whether there really is a man on the moon simply by showing how his presence there is possible.

This method of presupposing the unconditioned certainty of some propositional content, and thus of an unconditionally certain proposition, might perhaps allow the early Fichte's project to appear an interesting enterprise for the skeptic in some other respect, but it remains ultimately unconvincing with regard to the skeptic's fundamental objections. Fichte actually seems to have shared this view, even if this was not because he would have felt compelled to accept the skeptical criticisms of his approach. He could always insist that it is a misunderstanding to suppose that he is principally concerned with propositions or the conditions under which such propositions are known. He could point out, on the contrary, that the crucial thing for him is to show how we can explain knowing as a state of utter certainty, or, if we prefer to regard knowing as a capacity, to show which conditions one must introduce for the existence of a capacity for being utterly certain of something. Thus if Fichte nonetheless did find it necessary to question his own early approach, this may well have been because the objections of the skeptic had made it very clear to him that his own candidate for the utterly certain fact, namely a specific assertion or proposition, is too remote from the sphere which is supposed to be responsible for utter certainty, namely from the sphere of consciousness of the 'I' and its original 'Act'.²⁷

In any case, from 1796 until 1799, we find Fichte pursuing a different programme that is, fairly clearly, oriented once again to the skeptical challenge. This at least is suggested by the dictum from the *Second Introduction to the Wissenschaftslehre* (*Zweite Einleitung in die Wissenschaftslehre*) of 1797 that introduces this programme, and according to which: 'the derivation of an objective truth' is 'the sole purpose of all philosophy' (GA i/4. 210n.). If we may follow Fichte in characterizing the project of 1794/5 with the slogan 'from the fact to the Act', the new Fichtean programme is characterized for him by the declaration 'from the Act to the fact' (see, for example, the *Wissenschaftslehre according to a New Method* (*Wissenschaftslehre nova methodo*), GA iv/2. 33).

The governing idea of this new approach seems, once again, to be the claim that 'to demonstrate something' means 'to connect the truth of a proposition to another proposition. And thus, step by step, to arrive at a proposition, at a point A, which cannot be further demonstrated, and which may not be further demonstrated, for otherwise there is no truth whatsoever' (GA iv/2. 28). But, in contrast with the first version of the *Wissenschaftslehre*, this highest point is no longer identified through the analysis of something that 'is found in consciousness' (GA iv/2), but is now produced in a synthetic manner: 'He [the philosopher] must allow his I to advance SYNTHETICALLY,

²⁷ One of Fichte's remarks in the *Wissenschaftslehre in accordance with a New Method* clearly points in this direction. He observes that the non-I is here derived differently from the version of 1794/95, and characterizes the deficiency of this original version as follows: 'ABSOLUTE counter-positing is supposed to be demonstrated through the logical proposition -A not = A. Everyone will immediately concede this proposition in itself. But whence do I know that it is true? From experience? That is insufficient, for whence does experience know this?' (GA iv/2. 42).

allow his I to act before his eyes. Firstly to posit his I, and to observe it in its action in accordance with certain laws, and thereby to construct a world for himself' (GA iv/2).²⁸ Since Fichte characterizes this action of the I, and his observation of this action, through the concept of an 'Act' (GA iv/2. 9) which, *considered from a certain perspective* (acting at rest), produces a fact, a being, more precisely a concept (GA iv/2. 32–3), he can describe this procedure for discovering the highest point of all as a path that leads 'from the Act to the fact'.

There is no need at this point to examine how Fichte proceeds, from his concept of the Act, to construct the (real, existing) world and a (finite) intelligence that acts in this world, since this part of his new programme is not directly relevant to the problem of skepticism. It suffices to point out that Fichte's constructive procedure is based on the possibility of activating conceptual resources which, in Fichte's view, follow from the implications of the concept of an activity that can be intuited as such.²⁹ Thus 'determination' versus 'determinability', 'purposed' versus 'given', and 'I' versus 'non-I', present themselves as the central concepts here, whose subtle and various connotations are extremely difficult to pin down, and which Fichte exploits in a virtuosic and often obscure manner for his own constructive purposes.

If we ask whether Fichte's new approach can respond to the skeptical challenge more effectively than his original approach had done, the important aspect lies not in the constructive part that deals with the world and the human intellect, but in Fichte's starting point with an, or the, original 'Act'. For if it is correct that the programme of 1794/5 cannot unfold its full anti-skeptical power principally because the argument already presupposes the presence of an utterly certain fact of consciousness, thus proceeds from something 'given', a case of actual knowing, then it is evident that modifications must be introduced specifically with respect to this presupposition if there is to be any hope of countering skeptical objections. The most important of these modifications for Fichte is the recourse to what he designates with the concept of the original 'Act' as the only possible relevant point of departure. It is this concept therefore which must now assume the principal burden of anti-skeptical reflection and must therefore be made the centre of critical attention and examination.

One might think that Fichte's move from the fact to the original 'Act' as the relevant point of departure for the argument already has little chance of providing a rationally plausible or convincing alternative to the skeptic's position since the 'Act', just like the 'fact' in his earlier analysis, also enjoys the status of a presupposition. Fichte himself explicitly confirms this (see, for example, GA iv/2. 23). But if, as our previous reflections on the plausible objections to Fichte's earlier strategy suggested, it was precisely

²⁸ Although Fichte effectively criticizes his own original approach by redefining in this way his procedure for attaining the highest point, he explicitly blames Beck, rather than himself, for the programmatic error of attempting to discover a fundamental principle by analysis on the basis of a fact (and postulates). For Fichte's argument with Beck in this connection, see Klotz 2002, 23ff.

²⁹ On the methodological function of intuition in the *Wissenschaftslehre in accordance with a New Method*, see Klotz 2002, 44ff. See also Breazeale 1998, 587ff.

the procedure of beginning with a presupposition that led to the difficulties in refuting the skeptic, it would seem to make no difference whether we work with a fact or an Act as our presupposition. We should note, however, that the difficulties which were critically diagnosed in connection with presupposing a fact did not arise because there was a presupposition. These difficulties were connected with the problem that the presupposition had been proclaimed as a fact, as something simply found or given (see *GA* iv/2. 28). For it is solely this character of givenness which turns a fact into a problem if we are attempting to refute the skeptic since it is precisely with regard to anything found or given in experience that we can in principle be deceived in various respects.

Thus it is not immediately problematic that our starting point possesses the status of a presupposition. What is important, rather, is how the presupposition can be shown to be (a case of) knowledge that is entirely beyond dispute, that is invulnerable to any doubt, and to which we cannot possibly relate as something merely given.³⁰ But how does Fichte think that the introduction of the I as Act can help us any further in this connection? This leads us directly to the question what Fichte understands more precisely by this 'Act'—if it is to ground knowledge, how exactly is it to be described, and what defines its relevance for the process of knowing? According to Fichte, the Act can be described as a peculiar and unique activity of self-intuition. It is said to be unique insofar as it is a mental activity that is quite distinct from all other kinds of activity, and is specifically characterized by (1) reflexivity, (2) immediacy, and (3) productivity. The characteristic of reflexivity belongs to this activity by virtue of the object to which it is directed: it is supposed to be reflexive because it is expressly conceived as a kind of 'acting upon itself', i.e. it is supposed to be directed to that which it itself is, namely an activity that produces an 'I'. The characteristic of immediacy is ascribed to this activity by virtue of the conditions of its emergence: Fichte calls the activity 'immediate' or 'unconditioned' because it is directed towards nothing but its own enactment. And the characteristic of productivity points to the result of this activity: it is productive because it first produces what it is directed towards: namely the concept of the I.³¹ Fichte now presents the Act, so characterized, as a paradigmatic case of knowledge: the I can do no other than constitute itself in an original Act of knowing that concerns itself, and the thought (concept) 'I am' is nothing but the product of this knowing self-relation. If this is right, then anyone who can enact this I-constituting act already possesses some knowledge, namely a knowledge concerning itself, and this knowledge implies the reality, the being, of what is known. If we turn this reasoning against the skeptic, it indicates that we can demonstrate one indubitable case of knowledge which does not

³⁰ It should be noted that Fichte is not interested in claiming that what must be presupposed cannot *also* appear as a fact, as a given, but in showing that what is required as a presupposition cannot be regarded *merely* as a fact.

³¹ This characterization of the Act captures, in my view, the most important points to which Fichte refers on pp. 28–32 of the *Wissenschaftslehre in accordance with a New Method*. Dieter Henrich (1966) has already pointed out that this conception of the Act differs from that which he had developed in the first section of the *Wissenschaftslehre* of 1794/95. But these differences are of no consequence in the present context.

rely on anything given or presupposed, but rather, as an act that is constitutive of being, consists solely in the enactment of the Act itself.

To express the point in a way that is less dependent on Fichte's terminology, the idea of the I as an original Act is probably best captured by a dynamic model of description. On this kind of model, the factors which we wish to describe are presented by reference to the activities or powers (*Kräfte*) which constitute them. In this way the factors can be grasped as realizations of the characteristic configurations of elements which are required if we are to capture conceptually the activity or power in each case. On this descriptive model, the concept of 'Act' serves to characterize this, and solely this, specific dynamic structure. This structure is characterized, on the one hand, by an activity that is directed toward itself, while, on the other hand, it is directed towards the production of the elements that are necessarily required for the possibility of this activity. Fichte identifies this structure with what he calls the 'I', and identifies the activity which constitutes it with knowing.³²

It seems both probable and plausible to assume that with this description of the I in terms of activity or spontaneity, Fichte was taking up what he took to be a fundamentally significant aspect of Kant's conception of self-consciousness, understood as the transcendental unity of apperception, and attempting to develop this conception in a genuinely systematic fashion. For Kant's analysis of the 'I think' also ultimately seems to imply—especially in the second edition of the *Critique of Pure Reason*—that we must think of the I as a specific form of activity, and an activity which must be defined in terms of its synthetic accomplishments. But it is important for Kant's theory of the I that this activity can never become an object for itself, and that implies in turn that we cannot directly thematize the I independently of its products, that is, of its judgements.

³² This characterization is unfortunately hardly more illuminating than Fichte's own attempts to explicate his claims. It may therefore be helpful to point out that efforts to develop such a dynamic model are by no means peculiar to Fichte. Both Hegel and Schelling, e.g., also seem to favour this kind of model. The fundamental concepts of 'love' and 'life' in the early Hegel in particular, in the metaphysics of love and life that were sketched out in his pre-Jena period, can also probably be read in the context of this model. Thus 'love' is interpreted here as the expression of a complex relation which constitutes the 'self' and the 'other' through realizing an activity which only appears in the mode of a distancing appropriation and is stabilized in the form of 'being at home with oneself in the other', as Hegel puts it. The activity that Hegel here calls 'love' becomes a metaphysical principle as soon as one attempts (as the early Hegel for a short time did attempt) to interpret the entirety, or at least relevant dimensions, of reality by recourse to this structural character of this activity. And something similar appears to be true of Hegel's early concept of life: here too the concept of 'life' indicates a type of activity which Hegel regards as fundamental, one whose characteristic features can only be explicated by recourse to processes of relation and non-relation (separation) and the relation between the two. Similar examples of dynamic practical relations can also be found in the later Hegel, as in the model of self-consciousness developed by reference to the concept of desire in the *Phenomenology of Spirit*, or in the explication of the concept of 'friendship' in the 'Addition' to §6 of the *Philosophy of Right*.

Fichte's use of this kind of dynamic model to characterize the Act of the I allows him to introduce all the factors that are required for the possibility of this Act as necessary ones, such as 'subject' and 'object', for example. In introducing the concept of the I as Act, therefore, Fichte was not merely concerned with furnishing an appropriate description of the phenomenon, but also with addressing the methodological question concerning the correct, i.e. justified, introduction of factors which play a constitutive role in our understanding of reality.

According to Kant, what we call the 'I' (in the transcendental knowledge-constituting sense) is an empty representation that accompanies consciousness and is not as such a possible object of knowledge.³³

Now of course it hardly escaped Kant himself that his own doctrine of the I, on certain readings—and especially if we connect this doctrine with the thesis of the essentially unknowable character of the supersensible realm—could itself be interpreted as a position which either directly assumes a skeptical standpoint or which at least cannot avoid skeptical implications.³⁴ On these readings, the claim for the epistemic inaccessibility of the (transcendental) I, taken in conjunction with the thesis that its essential function is to make knowledge possible in the first place, implies the view that precisely what makes knowledge or cognition possible cannot itself be known or cognized. It is easy to see that any position which is compelled to endorse this view, and which is essentially interested in grounding the theory of knowledge, actually has no good way of countering skeptical objections: from the perspective of the skeptic, such a position grounds the possibility of knowing on something that cannot be known, and, to that extent, cannot properly be said to ground it at all.

For Fichte's concept of the I as an Act it is thus not only important to elucidate the internal character and structure of this Act, and to do so in some way or other in the 'spirit' of Kantian philosophy, if we are to realize the project of grounding knowledge in a manner that is immune to skepticism. In the context of this project, the principal question is rather the following: how can this Act be thematized in such a way that, on the one hand, it can furnish the basis for knowledge claims and, on the other, can be considered as epistemically accessible, or explained as an object of knowledge? We did not, of course, have to wait for the *Wissenschaftslehre according to a New Method* to see that Fichte's response to this question led directly to his theory of 'intellectual intuition'.³⁵ By introducing this concept of intellectual intuition Fichte was attempting to identify the character of that awareness which he claims is necessarily bound up with the idea of the I as an original Act.

³³ For more in this regard, see my chapter on the Paralogisms in Horstmann 1997, ch. 4.

³⁴ See Kant 1786, 324n. Kant, of course, regarded this interpretation as wholly mistaken (see the last sentence of the footnote).

³⁵ It is remarkable to observe, as has often been noted, that the concept of intellectual intuition plays no role at all in the *Wissenschaftslehre* of 1794/95, and indeed the term is not even mentioned there. And this despite the fact that in his review of *Aenesidemus* Fichte had already introduced intellectual intuition, even if only in passing and in the form of a undeveloped suggestion, as a reliable element for describing the self-relation of the I (GA I/3, 61). One reason for this remarkable circumstance may be this: insofar as the first version of the *Wissenschaftslehre* characteristically takes a fact as its point of departure, the question of how we are to acquire access to the original Act which is revealed as the condition of the possibility of this fact does not present itself as a particularly pressing one. For if we can explain the certainty which is ascribed to a fact as its characteristic feature only by recourse to such an Act, and if we are principally concerned with securing the possibility of this certainty, then the question of how this Act must present itself in accordance with its own intrinsic character is independent of the question concerning the possibility of certainty, and indeed subordinate to it. For Fichte's theory of intellectual intuition, see the still highly instructive study by Stolzenberg 1986, as well as Zöller 1998, 3ff. and Klotz 2002, 103ff.

The following consideration appears to have provided the basically straightforward reason behind Fichte's introduction of intellectual intuition: if we conceive of the I as an Act in the sense described, if we regard this conception as more than a purely theoretical construction, and attempt to reveal it as an intrinsically transparent Act that can be enacted by anyone, then we presuppose that this Act can somehow become present to us, can somehow be objectified. This claim is supposedly connected with the circumstance that each of us possesses the freedom to make ourselves, our own I, into an object of our thinking. For Fichte, the transparency and enactability of the act of the I reveal that the activity which is here considered as a knowing self-relation, and whose unique product can only be the concept of the I, in other words, the original Act, can only be enacted in the mode of conscious awareness: the enactment of this act of knowledge implies a conscious awareness of the act in question. Or to put the matter in another and probably rather abbreviated way: whenever I exercise the activity of thinking myself, I am aware of this activity of thinking myself and, to this extent, I am also immediately aware and conscious of myself.³⁶

Now the concept of intellectual intuition is supposed to characterize nothing other than the specific form of consciousness which is bound up with the very activity of thinking myself. There is no need here to investigate how Fichte came to adopt the term 'intellectual intuition', or how his own conception of it should be situated in relation to the discussion of his contemporaries concerning the possibility or significance of such a disputed faculty. The important thing is to grasp that Fichte's concept of intellectual intuition consists simply in the conscious awareness of the enactment of an activity which produces the concept of the I. It does not refer to the I as an object, whether it be empirically given or produced through intellectual construction, and thus has nothing to do, in Fichte's words, with a 'substance'.

This anti-skeptical strategy with respect to grounding knowledge, which proceeds from the Act that is present in intellectual intuition, does seem better suited for dispelling skeptical objections to the possibility of knowledge than the earlier approach based on a fact of consciousness. For now, at last, Fichte can show the skeptic that the philosopher does not have to presuppose anything that could be regarded in any incriminating sense as something simply as given, and further that the enactment of the Act which secures the possibility of knowing implies consciousness, i.e. the immediate presence or awareness of this enactment. The first point responds to the principal element of the skeptical criticisms directed at the initial version of the *Wissenschaftslehre*, while the second point responds to a skeptical objection that can be raised against the model of an original Act to which there seems to be no intelligible mode of access.

³⁶ This rather laborious description attempts to paraphrase a line of thought which Fichte presents in a highly condensed fashion, but which culminates in a formulation that is particularly relevant here: 'It is only by positing itself that the I becomes what it is—it does not already exist as SUBSTANCE—but to posit itself as positing is its essence, is indeed One and the same; it is therefore *immediately conscious of itself* (*ist sich seiner unmittelbar selbst bewusst*)' (GA iv/2, 31).

But even this new Fichtean approach may not entirely convince the stubborn skeptic—and not only because the skeptic could point out, once again, that the whole conception of an Act that is present to itself in intellectual intuition, as set forth principally in the *Wissenschaftslehre in accordance with a New Method* and in the *Second Introduction to the Wissenschaftslehre*, is still far too confusing, undeveloped, and obscure, for its anti-skeptical potential to be properly evaluated even as an overall approach. For this more or less practical objection must be further supplemented by a criticism which touches the substantive issue as well. According to this criticism, Fichte's new approach also ultimately involves various presuppositions which the skeptic is not bound to accept. The drift of the skeptic's argument can be sketched as follows: Fichte is indeed attempting to ground the possibility of knowledge; he hopes to accomplish this by demonstrating that the structure of the Act which produces the concept of the I necessarily leads to all the elements that are required for the possibility of knowledge in and about a given world. The possibility of knowledge is therefore strictly bound to the possibility of the activity of the I which posits its own being, its own reality. But an essential component of this view is the claim that we can only be aware of this activity in the mode of intellectual intuition because such intuition alone guarantees immediate certainty, that is to say, absolute incorrigibility and immunity to deception. But we must still ask the question: who, or what, is it that does the intellectual intuiting here? Must there not be some subject or other for this activity of intuition? In the context of Fichte's considerations, such a subject would have to be conceived as something that has no concept of itself, thus knows nothing of itself—it would be something resembling Sartre's pre-reflective cogito. But if it can have no concept of itself, and consequently knows nothing of itself, then it cannot be conceived as positing itself either, that is, cannot be conceived as existing, or as real. Yet that simply implies that we must presuppose the reality of the subject of intellectual intuition as factually given, as a fact. In the last analysis, therefore, even if we proceed from an original Act, we cannot avoid appealing to something that must be presupposed as a fact: if it was introduced in the *Wissenschaftslehre* of 1794/5 as the absolutely certain law of identity, it is now presented as the subject of intellectual intuition.³⁷ The I may well posit its own being by enacting the original Act, and to that extent it may be the case—as the skeptic can certainly concede—that it is also real insofar as it knows of itself, yet it is impossible to understand why this I is supposedly identical with the subject of that intellectual intuition which is alleged to accompany this knowledge of itself on the part of the I. But if we cannot claim such an identity, then the reality of this subject is merely presupposed, i.e. assumed as factually given, or—in Fichte's terminology—we find that being is here introduced as a fact of consciousness. And the skeptic is perfectly free to challenge this fact. To that extent, this anti-skeptical strategy, which takes its point of departure from the original Act, must also be considered a failure. For Fichte

³⁷ For further consideration of this problem, see Klotz 2002, 112. Klotz himself believes that Fichte is capable of meeting the objection.

this failure seems to have been a sufficient reason for abandoning entirely any direct engagement with the skeptical challenge, and with it any approach to the theory of knowledge that is essentially oriented to the question of justification.

V

Fichte's great project is to present a theory which shows that what we rightly say we know, must also be the case, must also be real. For it is only if we can show this, according to Fichte, that we are in a position to demonstrate that the Kantian distinction between appearances and things in themselves is epistemically irrelevant or superfluous, and thus, on the one hand, avoid a fundamental problem supposedly afflicting Kant's theory of knowledge, and, on the other, to buttress the claim that there is indeed some human knowledge that is immune to all skeptical doubt. As we have seen, Fichte initially seems to have hoped, until about 1800 at least, that it was possible to accomplish this project in a manner that still remained ultimately indebted to a philosophical model oriented to the issue of justification, i.e. one that undertook to show that there are some factors or features that are, so to speak, self-verifying in character. If these factors are expressed as judgements (propositions), then the following claim is supposed to hold: if they are 'certain' (which for Fichte is equivalent to 'known'), then they are also true (which is the same for Fichte as 'possessing real content' or being 'the case'); and, conversely, if they are true, then they are also certain. Fichte identifies the activity which he dubs the 'Act' as just such a factor or feature, an activity which—in good Cartesian fashion—he coordinates with the proposition 'I am' (in one or other of the various formulations he chooses to deploy) as an equivalent expression. If it is likely that, even after 1800, Fichte continued to believe that with his concept of the original Act he had, substantively speaking, identified an indubitable foundation for the justification of our claims to knowledge, it is nonetheless clear, for what I hope are now sufficiently transparent reasons, that his introduction of this concept, whether in the manner of the *Wissenschaftslehre* of 1794/5 or in that of the subsequent *Wissenschaftslehre in accordance with a New Method*, would still not necessarily or inevitably convince the resolute skeptic. In the last analysis, Fichte would have to confess to himself, any appeal to some privileged factor or feature which finds expression in a self-verifying judgement is already condemned to failure on methodological grounds: one cannot avoid presupposing something that the skeptic can always properly challenge. Once he acknowledged this, Fichte realized that we must abandon the attempt to furnish, in any version, a justification-oriented refutation of the skeptic based upon an appeal to some allegedly indubitable principle or proposition.

The various versions of the *Wissenschaftslehre* which Fichte produced after 1800 represent a new attempt, or perhaps several new attempts, to accomplish his great project of grounding knowledge against all skeptical claims in a way that, at least in his eyes, was much more convincing. These versions arose at different stages of his subsequent

career and (with a single exception) were never published by Fichte himself. They resemble one another not only in the fact that they are all highly obscure and difficult to penetrate, but also, and principally, in the fact that they no longer place the search for, or the plausible legitimization of, some privileged factor or feature in the foreground of attention. On the contrary, they pursue what, in a programmatic formulation from the *Wissenschaftslehre* of 1801/2, Fichte calls 'the analysis of absolute knowing' (GA ii/6. 274) and make this the starting point and the central issue of his subsequent reflections. In these versions of the *Wissenschaftslehre* the direct engagement with skepticism retreats into the background in the sense that the refutation of skepticism is no longer explicitly proclaimed as a basic motivation for the philosophical reflections which follow. Fichte now asserts, instead, that his investigations are principally addressed to the question concerning the proper grounding of knowledge, i.e. the elucidation of the conditions which must be fulfilled whenever we claim to really know something. For reasons of space, the following remarks will concentrate upon the *Wissenschaftslehre* of 1801/2 since I am principally interested here in indicating and describing the new approach adopted in this text, rather than in trying to evaluate Fichte's various attempts to present and justify his new approach in systematic form.³⁸

It is already symptomatic of this change of priorities—the relative displacement of the skeptical question regarding justification and the new emphasis upon analysing the concept of knowledge—that the necessity of presuppositions and the appeal to facts no longer appear as taboo in Fichte's eyes. For he now starts directly from a fact, namely the fact that there is knowledge (see GA ii/6. 137, 141, 206), and that there is a subject that knows, a consciousness, a thinking I (see GA ii/6. 130–1), and proceeds to ask what we should properly understand by knowing, that is to say, how we can describe the concept of knowing. The method which he adopts for the purpose of this description is oriented to a kind of procedure that rather resembles what we would now associate with phenomenological investigations.³⁹

According to this approach, knowing something is a mental state which is characterized by an utter certainty that is indicated by the 'constancy, stability and unshakability of our representing' (*Stetigkeit, Festigkeit, und Unerschütterlichkeit des Vorstellens*) (GA ii/6. 136). This mental state obtains when it is utterly impossible for me to think that what I claim to know could ever be otherwise. But this impossibility is only given where the knowing subject can enjoy a comprehensive view (*Übersicht*) of all the conceivable situations which are relevant to the knowledge claim being made, and can exclude an alternative claim for every one of these situations. This comprehensive view cannot, for Fichte, result from a process which the subject actively runs through, as if, for example, the subject who claims to determine the sum of the angles of the triangle

³⁸ For more on the way in which the later versions of the *Wissenschaftslehre* define their distinctive tasks, see Schlösser 2006.

³⁹ According to Zöllner 1998, 33, this phenomenological mode of procedure is also characteristic of the Fichte of the *Wissenschaftslehre in accordance with a New Method*.

would first have to examine any number of triangles one after another. It is rather the result of an instantaneous grasp of all possible cases, 'the grasp of a certain representing [Fichte is referring to the representing of what we claim to know] . . . in its totality at an utterly single glance' (*in seiner Gesamtheit schlechthin mit Einem Blick*) (GA ii/6. 137–8), or an 'absolute synoptic grasp and comprehensive view of a manifold of representations' (*absolutes Zusammenfassen und Uebersehen eines Mannigfaltigen von Vorstellungen*) (GA ii/6. 138). Fichte describes this grasp and this view of a manifold at a glance as 'intuiting' or 'beholding' (*Anschauung*). Knowing, for him, is therefore intuition (GA ii/6. 138; see also p. 145).

Now what holds for any claim to knowledge in relation to any object whatsoever—Fichte calls this 'relative' (GA ii/6. 144) or 'particular knowing' (GA ii/6. 145)—or what is therefore common to all particular, relative knowledge, this, according to Fichte, constitutes what must be described as 'absolute knowing'. Absolute knowing exhibits the 'character of knowing in general which any particular knowing must possess', and must be 'presupposed for the possibility of every specific, particular knowing'. In other words it is the 'absolutely One and the same in all particular knowing' (GA ii/6. 146). Although this designates what is common to all particular knowing, it cannot be grasped, according to Fichte, as the product of an abstraction ('which is supposed to elicit from a plurality of individuals something that does not lie in any single one of them', GA ii/6. 146). This absolute knowing is a constitutive condition for relative knowing—for without participation in absolute knowing, no relative knowing is possible either.

If we bear these reflections in mind, it is clear why Fichte claims that the analysis of absolute knowing is a condition of any theory of knowing whatsoever. It is far less clear, on the other hand, how precisely he pursues this analysis. He starts from the concept of the absolute since the concept of knowing that interests him is indeed supposed to be one of absolute knowing. For Fichte, it must hold for the absolute that it cannot be further characterized at all, that it is a whole without predicates, precisely because 'any further word that we may add to the expression "the absolute" would destroy "absoluteness as such" and transform it into something that is still qualified . . . by the perspective and relation of the word we have added' (GA ii/6. 143). The absolute is the utterly ineffable.⁴⁰ Fichte holds that this absolute, which cannot be conceptually determined in any way, is conceivable for us (i.e. is susceptible to tentative description) in terms of two aspects: 'on the one hand, that it is utterly *what* it is, that it rests on and in itself, without any change or alteration, is firm, complete and self-enclosed, and, on the other, that it is *what* it is simply because it is, from itself and through itself, without any external influence, insofar as alongside the absolute there remains nothing external whatsoever, so that everything that is not the absolute itself vanishes here' (GA ii/6. 143).

⁴⁰ For Schlösser 2004, 310ff., this conception expresses a kind of modified Spinozism which Fichte developed under the direct influence of Jacobi.

This way of introducing the concept of the absolute seems to be an attempt to respond to an objection that had already played a role in the controversies surrounding Fichte's first version of the *Wissenschaftslehre*.⁴¹ The principal motivation for this critical objection then was the suspicion that Fichte's early approach, which presented the absolute I as ultimately responsible for positing all being, was compelled to treat reality as a product of thought, as something that could only be regarded as the result of a conscious activity and could never even be conceived as anything but thought. Against this view, it was insisted that the discursive character that necessarily belongs to thought must always bind thinking to the form of the judgement, that is, to a separation between subject and predicate. Thus a purely discursive concept of reality, i.e. one in which reality is already conceptually or intellectually structured in advance, is not in a position properly or adequately to reveal that originally unified and undifferentiated character of reality which is suggested by expressions such as 'being simply as and because it is' or is broadly captured by Sartre's talk of reality as the 'in-itself'. Yet our discursive access to reality, mediated as it is by the form of judgement, presupposes that what presents itself in the subject-predicate judgement as separate, really forms an undifferentiated unity since judgement itself is precisely the separation of the originally undifferentiated into subject and predicate.⁴² If we wish, therefore, to develop an adequate concept of reality, we must acknowledge and do justice to this essential dimension of undifferentiated unity which precedes all judgement and makes it possible in the first place. But this, it was complained, is precisely what no discursive conception of reality can accomplish: it cannot grasp reality as 'primordial being' or 'being prior to thought' (*unvordenkliches Sein*) (Hölderlin), and it cannot do so because it has entirely overlooked that judging (the conceptual) presupposes being (the non-conceptual).

Now Fichte deploys his concept of the absolute to characterize what he calls 'absolute knowing'. To remind ourselves: absolute knowing is not supposed to signify the knowing of something specific, and is not therefore a special case of relative or particular

⁴¹ This objection can be traced back to critical reflections which appear to have originated with Hölderlin and his circle in Bad Homburg (Sinclair, Zwilling, and others). See Henrich 1965/6 and Henrich and Jamme 1986.

⁴² This conception of judgement as an 'original division' (of a whole into a subject-predicate relation) is principally associated with the name of Hölderlin in the context of early idealist philosophy. The basic idea behind this approach appears to be the following: the judgement somehow 'tears apart' what is presented here as a single or undivided state of affairs. Thus e.g. the judgement 'the rose is red' is the discursive expression of what is given in unified intuition as a red rose. The object of this unified intuition is represented as something that is not itself conceptually structured, although it is possible to relate to this object conceptually through the judgement. And this supposedly signifies that the judgement is related to something (the state of affairs it expresses) that undividedly contains what can only appear in the judgement as divided, namely as the 'subject' to which the 'predicate' is ascribed. To this extent, so it seems to this critical perspective, the conceptual or discursive presentation of the state of affairs is defective insofar as it cannot express the original, unified, and undifferentiated character of the state of affairs in question, even though this undifferentiated character defines the latter as it 'actually' is. On this conception, therefore, the judgement cannot, as Hegel later puts it, present the proper 'form of truth', i.e. cannot furnish the appropriate medium for presenting what something 'in truth' really is.

knowing, and is not supposed to be understood as a product of abstraction from what all particular cases of knowing have in common. To this extent, Fichte's absolute knowing is regarded as opposed to the knowing of something (see ii/6. 145). Since all knowing is intuition, and absolute knowing is a knowing that has no object-term, no intentional object—for it is precisely not supposed to be a knowing of anything specific—then absolute knowing, according to Fichte, can only be described as an intuiting that grasps itself. If we now ask what characterizes such intuition, Fichte's remarks concerning the absolute start to play a role. Absolute knowing, according to Fichte, is supposed to be absolute, that is, it is supposed to reveal itself as 'One, precisely as *knowing* that is the same as itself and remains eternally the same (*sich selbst gleiches, und ewig gleichbleibendes*) as the unity of the single and indeed highest intuition' (ii/6. 147). Because this knowing is absolute, it must exhibit the characteristics of the absolute: it must be an 'actual *real* unification' (*Vereinigung*) (ii/6. 150) of these characteristics. Since Fichte has designated the latter by recourse to the concepts of 'being' and 'freedom', he now understands absolute knowing to mean the real unification of being and freedom. And since absolute knowing is supposed to be the ground of the possibility of all particular knowing, and thus transcends all such knowing, he can furnish the following description or 'real-explanation' (*Real-Erklärung*), as he calls it: 'Beyond all knowing [here Fichte means particular knowing] . . . freedom and being come together, interpenetrate with one another, and this innermost penetration and identification of both into a new form of being (*Wesen*) now yields knowing for the first time precisely as knowing, as an absolute Thus (*Tale*)' (GA ii/6. 148). And Fichte adds: 'Everything depends on insight into this point.'

According to Fichte, this mutual interpenetration of being and freedom in knowing is supposed to possess the 'formal' character of 'being for itself'. He surely means that this interpenetration is an event that, once it transpires, opens up or releases a consciousness or knowing of itself (the event, i.e. the interpenetration). It is at this point that Fichte introduces his image, as famous as it is notorious, of the 'eye' that cannot properly be said to accompany or observe this interpenetration, but is rather a constitutive moment of the latter: absolute knowing is essentially an interpenetration which is enacted in the mode of a conscious self-present awareness of that which interpenetrates (*eine Durchdringung, deren Vollzug selbst im Modus der bewussten Präsenz dessen, was sich durchdringt*). In this connection Fichte specifically indicates that he has designated 'this absolute self-penetrating through itself and being for itself' (GA ii/6. 150) with the concept of 'I-hood' (*Ichheit*), and regrets, not indeed without good reason, that what he means by this concept is obviously difficult to grasp. And this not least because Fichte does not understand his exposition of what he calls 'absolute knowing' as a discursive argument, but proclaims it rather as the result of an inner intuition which only those who are not 'inwardly blind' (GA ii/6. 150) are capable of enacting.

Whatever we may think of this intuition which we are expected to enact for ourselves, one to which Fichte's exposition of absolute knowing directly appeals, we can

also at least attempt to express the underlying idea without resorting to the language of intuition. For let us think of reality in its entirety, the absolute, as a completely undifferentiated unity, of which all we can say is that it is. Let us think, further, that this 'One', which almost recalls the 'Being' of Parmenides, since it is everything, contains subjects who can consciously turn aspects of this reality into objects for themselves, i.e. who are capable of relating to some aspect or other of this reality in the mode of knowledge or cognition. Since these knowing and cognizing subjects are part of reality, they cannot, if genuine insight into this reality is to be ascribed to them, be conceived as if they would have to relate epistemically to reality from a standpoint that lay, so to speak, 'outside of reality', as if they could behold and describe reality 'from outside'. For there is no such standpoint. To express this in the language of Fichte's time: the opposition between the (knowing and cognizing) subject and the (known, cognized) object cannot be 'absolute' if knowing is to be possible at all. These subjects must therefore be conceived as capable of relating epistemically to reality 'from within', as integrated within reality. And the relation in question must be such that it allows us to maintain a double difference: on the one hand, it must be possible to distinguish between reality as an existing undifferentiated unity (as something like Sartre's 'being in itself') and reality as an object of knowledge. For if it were impossible to make this distinction, we could never secure the presupposition that our knowledge claims have any relation to reality. On the other hand, we must be able to grasp this relation as that of a subject to something other than itself. For otherwise it would be impossible to understand that knowledge claims must involve more than simply claims concerning the inner states of a subject. If we acknowledge this double restriction, one can see why the possibility of knowing, understood as a relation between a subject and an object, arguably depends upon a creative conceptual act that produces the subject on one side and a manifoldly differentiated reality on the other. For a conscious or knowing relation to reality is only possible for the subject if the latter can grasp reality 'in itself', containing the subject and everything else in undistinguished indeterminacy, as an internally differentiated structure constituted in such a way that the subject can distinguish between itself and what is other than itself. The subject must therefore be able to conceive itself as something 'for' which something is, and that which is 'for' the subject must possess the character of reality. Such a self-conception is the necessary condition for the possibility of speaking of any relation of knowing between a subject and some actual state of affairs. Since this self-conception involves an achievement, an activity, it can be regarded as a creative or, in Fichtean language, 'positing' act. The act in question is not produced or occasioned by anything else, and thus springs from a kind of free decision. It creates, at one and the same time, both the subject and a world that is differentiated for the subject.⁴³ The subject that is self-created in this sense is nothing further and nothing more than the subject of knowing—and it is nothing at all independently of, or

⁴³ By 'world' here I understand 'reality in itself from the perspective of the conscious subject'. I have chosen this terminology by drawing on Heidegger's language in *The Age of the World Picture*.

outside of, this relation of knowing. And that is because it invents or generates itself, so to speak, as the element of that structure of knowing 'for' which something can exist in the mode of knowledge, and apart from this mode of knowing it has no function whatsoever. For this subject, the world to which it relates through knowing is a unitary reality, no longer an undifferentiated unity 'in itself', but rather a unity which is present as the unity (for Fichte: as the absolute interpenetration) of what is conceptually distinguished (for Fichte: of being and freedom). To be conscious of reality, or to know about reality, is thus intrinsically bound to a complex act which conceptually generates both self and world, or, in Fichte's language, which posits both self and world. This act does not indeed create reality, but it creates a form in which reality appears, namely the form of knowing or cognition.

I fear that this attempt to furnish a relatively brief elucidation of what I have here presented as Fichte's governing philosophical idea may seem hardly more intelligible than the clarifications which he himself provided. And, as is well known, they were never that clear or illuminating. But in any case, to return to Fichte directly, it certainly makes sense, in the light of his arguments, to claim two things. In the first place, his analysis does indeed yield a thesis which, if taken out of context, can lead (and certainly has led) to radical misunderstandings. One need only consider statements like the following: 'there is no being whatsoever (namely for knowing, for this view rests entirely within the standpoint of the latter) other than through knowing itself' (GA ii/6. 159; see also p. 272). But one must simply recognize that Fichte's claim here has nothing to do with any material creation on the part of the knowing subject since he is deploying the term 'being' here as a primary category, or fundamental concept, without which the discursive and conceptually structuring constitution of reality that is necessary for knowledge would be impossible. In the second place, given his own presuppositions, Fichte must insist that 'the focus of all actual knowing = the I' (GA ii/6. 275). He is not adverting here to the trivial point that there must always be some subject of knowing, someone to whom knowledge claims may be ascribed. He is rather expressing the insight that the possibility of grasping reality as knowable requires the conceptually structuring achievement of an activity which, by accomplishing this achievement, conceives or posits itself in accordance with the conceptual conditions which transform reality 'in itself' into a knowable world.

How then would the skeptic respond to all this? If he considers Fichte's reflections seriously, would he still have grounds to doubt the possibility of knowledge? One thing seems to be clear: the kind of objections which the skeptic raised against the way in which Fichte pursued his original anti-skeptical programme both in the *Wissenschaftslehre* of 1794/5 and in the *Wissenschaftslehre in accordance with a New Method*, are not particularly effective against the approach which Fichte adopted in 1801/2. And this is because with this new approach Fichte has finally abandoned the idea of trying to justify knowledge demonstratively, and thus has ceased to privilege a procedure which is ultimately modelled on the demands of a justification-oriented method of refuting skepticism. One may well feel that Fichte's new approach from 1800

onwards reflects the insight that any attempted justification of knowledge cannot avoid appealing to some facts the acceptance of which cannot itself be justified, and which must be presupposed as *facta bruta* which the skeptic remains at liberty to doubt. To justify is to furnish grounds, and for these grounds there are further grounds in turn, which leads us either to an infinite regress, or to a point where we must admit that they cannot be further grounded. And this calls forth the skeptic who will now insist that whatever cannot be demonstratively grounded can always, in principle, be suspected of error.

The project of 1801/2 pays no attention whatsoever to the earlier strategy for countering skepticism which was doomed to fail from the first. Fichte now develops a theory according to which knowing makes a relation between the I and the world possible in the first place. For knowing is here understood as an act which precedes the distinction between the I and the world in the sense that it constitutes something like a conscious (knowing) subject and a conceptually structured reality, i.e. a world. Knowing is the fundamental enactment of a structuring achievement in which the I and the world are founded. This fundamental act may or may not transpire (GA ii/6. 214), but it must transpire if there is ever to be anything real for a subject. One could formulate this in the phrase: only through knowing is there an I and a world. This does not mean, of course, that in some imaginary place or other there exists some metaphysical substrate called 'knowing' which then somehow 'creates' the I and the world. What Fichte means is that only where there is the mode of knowing, understood as a kind of original comportment, is there also a conscious I, an I that knows of itself, and a knowable world. If this is so, the skeptic will find it difficult to convince us with his general suspicion concerning the possibility of error. The skeptic may indeed put in question everything that belongs to what Fichte calls 'relative' or 'particular' knowledge—and one may always actually err in relation to any specific concrete claim to knowledge—but he will not be able to deny that the attitude of 'knowing' which we adopt towards states of affairs is not only possible, but also necessary. And this because without the existence or reality of this attitude there is not only no subject, and no world distinct from the subject, but also because without knowing in general no (conscious) attitude or comportment, of whatever kind to whatever object, and therefore no doubt, would even be conceivable.

Now one might object that anti-skeptical Fichtean reflections of this kind are obviously operating with a concept of knowing which is quite different from our current concept of knowledge, insofar as the former robs the standard modern concept of its basic intentional structure, of its essential relation to something specific in each case. What we mean by 'knowledge', according to this objection, is really always what Fichte characterizes as 'particular knowing', and when we ignore or abstract from this idea of knowing, as Fichte does, we distort or do violence to the only sense of knowledge that is relevant to the skeptic. But we should not forget that Fichte would insist that (1) his own concept of (absolute) knowing underlies all particular claims to knowledge, and that (2) the skeptic, if he understands what he is about, principally directs his doubts not against our 'particular' knowledge claims, but rather casts doubt on knowing itself

as a possible attitude or posture towards reality, one that is distinct from other mental attitudes such as 'believing' or 'opining'. If this is right, then one cannot simply appeal to our current concept of knowing as a decisive objection to Fichte.

Of course, one would now have to ask how Fichte develops his own approach to grounding in such a way as to yield a relatively standard or 'normal' theory of our epistemic relation to the world, one that would explain why we rightly believe that in our everyday epistemic comportments we have to do with the world as we have actually experienced it. But here is not the place to pursue this question any further. For in the last analysis we are concerned only with the anti-skeptical potential of Fichte's extremely radical approach to philosophy. Although we certainly cannot claim that Fichte's procedure for establishing a concept of knowing that is immune to skeptical doubt, namely his insistence upon (inner) intuition alone as the only legitimate access to this concept, presents an immediately convincing one, and although we cannot claim that Fichte's proposal for analysing knowing by reference to the categories of 'being' and 'freedom' is immediately illuminating either, he nonetheless succeeds remarkably well in doing justice to a very deeply rooted intuition that is quite independent of his own philosophy. For if we pursue Fichte's approach, we can see very clearly why, once we have left 'the study' (as Hume puts it), it is so difficult to take the skeptical challenge to the possibility of knowledge with any real seriousness. If knowing, understood in Fichte's sense, is actually the condition under which both we ourselves as conscious beings and reality as a knowable world is given for us, then to doubt the possibility of knowing, i.e. to take the skeptic seriously, would imply that we could somehow inhabit a realm beyond and outside of ourselves and our world. If we follow Fichte's argument, then the very idea of such a standpoint is not only unserious, but manifestly absurd. It is another question, of course, whether his own path to this insight, which he would have us travel too, is really open to us.⁴⁴

Translated by Nicholas Walker

References

- Breazeale, Daniel (1980–1), 'Fichte's "Aenesidemus" Review and the Transformation of German Idealism', *Review of Metaphysics*, 34, 545–68.
- Breazeale, Daniel (1982), 'Between Kant and Fichte: K. L. Reinhold's "Elementary Philosophy"', *Review of Metaphysics*, 35, 785–821.
- Breazeale, Daniel (1998), 'Fichte's *Nova Methodo Phenomenologica*: On the Methodological Role of "Intellectual Intuition" in the Later Jena *Wissenschaftslehre*', *Revue Internationale de Philosophie*, 4, 587–616.

⁴⁴ I should like to express my thanks to Dina Emundts and Ulrich Schlösser for helpful critical discussion of these issues and for drawing my attention to various sources and contributions.

- Breazeale, Daniel (2003), 'Der Satz der Bestimmbarkeit: Fichte's Reception and Transformation of Maimon's Principle of Synthetic Thinking', *Internationales Jahrbuch des Deutschen Idealismus*, 1, 115–40.
- Di Giovanni, George, and H. S. Harris (eds) (2000), *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism* (Indianapolis: Hackett) [contains selections from K. L. Reinhold, *The Foundation of Philosophical Knowledge*, and G. E. Schulze, *Aenesidemus*, and Fichte's *Review of Aenesidemus* (complete)].
- Fichte, Johann Gottlieb (1964–), *Gesamtausgabe der Bayerischen Akademie der Wissenschaften*, ed. Reinhard Lauth et al. (Stuttgart-Bad Cannstadt: Frommann-Holzboog) [cited as GA, followed by section, volume, and page number].
- Fichte, Johann Gottlieb (1982), *Science of Knowledge*, ed. and tr. Peter Heath and John Lachs (Cambridge: Cambridge University Press) [contains Fichte's *Foundations of the Entire Wissenschaftslehre* (*Grundlage der gesamten Wissenschaftslehre* 1794/95)].
- Fichte, Johann Gottlieb (1988), *Early Philosophical Writings*, ed. and tr. Daniel Breazeale (Ithaca, NY: Cornell University Press) [contains Fichte's *Review of Aenesidemus* and *Concerning the Concept of the Wissenschaftslehre*].
- Franks, Paul (2003), 'Does Post-Kantian Skepticism Exist?', *Internationales Jahrbuch des deutschen Idealismus*, 1, 141–64.
- Henrich, Dieter (1965/66), 'Hölderlin über Urteil und Sein', *Hölderlin-Jahrbuch*, 14, 73–96. English translation: 'Hölderlin on Judgement and Being', in Dieter Henrich, *The Course of Remembrance and Other Essays on Hölderlin* (Stanford, CA: Stanford University Press, 1997), 71–89.
- Henrich, Dieter (1966), 'Fichtes ursprüngliche Einsicht', in Dieter Henrich and Hans Wagner (eds), *Subjektivität und Metaphysik* (Frankfurt am Main: Klostermann), 188–232. English translation: 'Fichte's Original Insight', tr. David Lachterman, in *Contemporary German Philosophy*, i (Philadelphia: Pennsylvania State University Press, 1982), 15–53.
- Henrich, Dieter, and Christoph Jamme (eds) (1986), *Jakob Zwillings Nachlaß. Eine Rekonstruktion* (Bonn: Bouvier).
- Horstmann, Rolf-Peter (1997), *Bausteine kritischer Philosophie. Arbeiten zu Kant* (Bodenheim: Philo Verlagsgesellschaft).
- Horstmann, Rolf-Peter (2000), 'The Early Philosophy of Fichte and Schelling', in Karl Ameriks (ed.), *The Cambridge Companion to German Idealism* (Cambridge: Cambridge University Press), 117–40.
- Horstmann, Rolf-Peter (2004), *Die Grenzen der Vernunft. Eine Untersuchung zu Zielen und Motiven des deutschen Idealismus* (Frankfurt: Klostermann).
- Jacobi, Friedrich Heinrich (1787), *David Hume über den Glauben oder Idealismus und Realismus. Ein Gespräch* (Breslau). English translation: *David Hume on Faith, or Idealism and Realism: A Dialogue*, in F. H. Jacobi, *The Main Philosophical Writings*, tr. and ed. George di Giovanni (Montreal and Kingston: McGill-Queen's University Press), 253–338.
- Jacobi, Friedrich Heinrich (1789), *Über die Lehre des Spinoza in Briefen an Herrn Moses Mendelssohn*, Breslau. Partial English translation: *Concerning the Doctrine of Spinoza in Letters to Herr Moses Mendelssohn* (1789), in F. H. Jacobi, *The Main Philosophical Writings*, tr. and ed. George di Giovanni (Montreal and Kingston: McGill-Queen's University Press), 339–78.
- Kant, Immanuel (1786), 'Was heißt: Sich im Denken orientieren?', in *Berlinische Monatsschrift*, 304–30. English translation: 'What is Orientation in Thinking?', tr. H. B. Nisbet, in Kant, *Political Writings* (Cambridge: Cambridge University Press, 1991), 237–49.

- Klotz, Christian (2002), *Selbstbewußtsein und praktische Identität. Eine Untersuchung über Fichtes Wissenschaftslehre nova methodo* (Frankfurt: Klostermann).
- Kompa, Nikola (2001), *Wissen und Kontext: Eine kontextualistische Wissenstheorie* (Paderborn: Mentis).
- Lauth, Reinhard (1989), 'Die Entstehung von Fichtes Grundlagen der gesamten Wissenschaftslehre nach den Eignen Meditationen über Elementar-Philosophie', in *Transzendente Entwicklungslinien von Descartes bis zu Marx und Dostojewski* (Hamburg: Meiner).
- Reinhold, Karl Leonhard (1789), *Versuch einer neuen Theorie des menschlichen Vorstellungsvermögens* (Prague/Jena: C. Widtmann & I. M. Mauke).
- Schlösser, Ulrich (2004), 'Modifikationen des Spinozismus', in Walter Jaeschke and Birgit Sandkaulen (eds), *Friedrich Heinrich Jacobi. Ein Wendepunkt der geistigen Bildung der Zeit* (Hamburg).
- Schlösser, Ulrich (2006), 'Worum geht es in der späteren Wissenschaftslehre und inwiefern unterscheiden sich die verschiedenen Darstellungen von ihr dem Ansatz nach?', *Fichte-Studien*, 30, 15–23.
- Schulze, Gottlob Ernst (1792), *Aenesidemus oder über die Fundamente der von dem Herrn Prof. Reinhold gelieferten Elementar-Philosophie. Nebts einer Vertheidigung des Skeptizismus gegen die Anmaßungen der Vernunftkritik* (Helmstädt). Partial English translation in di Giovanni and Harris 2000, 104–33.
- Stolzenberg, Jürgen (1986), *Fichtes Begriff der intellektuellen Anschauung. Die Entwicklung in den Wissenschaftslehren 1793/94 bis 1801/02* (Stuttgart: Klett-Cotta).
- Van der Zande, Jan, and Richard H. Popkin (eds) (1998), *The Skeptical Tradition around 1800: Skepticism in Philosophy, Science and Society* (Dordrecht: Kluwer).
- Vieweg, Klaus (1999), *Philosophie des Remis. Der junge Hegel und das 'Gespenst des Skepticismus'* (Munich: Fink Wilhelm).
- Zöller, Günter (1998), *Fichte's Transcendental Philosophy: The Original Duplicity of Intelligence and Will* (Cambridge: Cambridge University Press).

6

Fichte's Transcendental Ethics

Paul Guyer

1. The Challenges for a Transcendental Moral Philosophy

Johann Gottlieb Fichte's *System of Ethics* (*Das System der Sittenlehre nach den Principien der Wissenschaftslehre*), published in 1798,¹ is, considered with regard to Kant, two or three books for the price of one. In its first two parts, it deals with the foundational issues for practical philosophy that Kant had dealt with in the *Groundwork for the Metaphysics of Morals* of 1785 and the *Critique of Practical Reason* of 1788, while in its third part it deals with the duties of virtue that Kant had dealt with in the second half of his *Metaphysics of Morals*, the Doctrine of Virtue, first published separately in August 1797, and then unified with the first half of the work, the Doctrine of Right, in 1798, thus the same year as Fichte's work. (Fichte dealt with the subject matter of the Doctrine of Right, namely legal and political obligations, in his *Foundations of Natural Right* of 1796, thus before he could have seen Kant's Doctrine of Right, and it is quite likely that his *System of Ethics* was also conceived and largely drafted before he could have seen Kant's Doctrine of Virtue.) But Fichte's *System of Ethics* does not merely cover the same territory as Kant's foundational works in practical philosophy on the one hand and more specific or applied theory of ethical duties on the other: his work is also clearly an attempt to improve upon Kant's foundations for practical philosophy, to use the method of transcendental philosophy as he conceives it to solve outstanding problems in Kant's own attempt to ground moral philosophy within the framework of the transcendental philosophy. Examining the success of Fichte's attempt at a transcendental moral philosophy is thus a way at least to begin considering the prospects for such an enterprise beyond Kant's own obviously flawed attempt at it.²

¹ Johann Gottlieb Fichte, *The System of Ethics According to the Principles of the Wissenschaftslehre*, ed. and tr. Daniel Breazeale and Günter Zöller (Cambridge: Cambridge University Press, 2005). Cited as SE.

² For my diagnosis of the flaws in Kant's own attempt at a transcendental practical philosophy, see my 'Naturalistic and Transcendental Moments in Kant's Moral Philosophy', *Inquiry* 50 (2007), 444–64.

Here are three outstanding problems in Kant's foundations for practical philosophy that Fichte can be thought of as attempting to resolve by the use of the transcendental method of philosophy as he understands it in the *System of Ethics*. First, Kant clearly had difficulties with the issue of a transcendental deduction of the validity of the fundamental principle of morality for human beings, that is, with the deduction of the categorical imperative (the former presents itself to us as the latter, of course, for even if the fundamental principle of morality is valid for human beings it does not exhaust our possible motivations, but rather presents itself to us as a constraint on at least some of our desires, thus as a categorical imperative, 'the relation of an objective law of reason to a will that by its subjective constitution is not necessarily determined by it (a necessitation)'):³ in section III of the *Groundwork*, Kant attempted to demonstrate that the moral law is the law for our authentic selves by arguing that the moral law is necessary for any rational being and that the uniqueness of our faculty of reason shows that at the noumenal level we are rational beings (*G* 4: 451–3; Gregor, 98–9), an argument that obviously depends upon a transcendent claim about our noumenal selves rather than any merely transcendental analysis of or argument about the conditions of the possibility of our phenomenal consciousness; and then in the *Critique of Practical Reason* Kant obviously recoiled from this relapse into transcendent metaphysics, insisting instead that all human beings are immediately aware of the obligatory status of the moral law as soon as they attempt to reason their way through any moral issue (*CPracR* 5: 30; Gregor, 163–4), a 'fact of reason' that cannot be reasoned 'out from antecedent data of reason' but that 'forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition' (*CPracR* 5: 31; Gregor, 164) and which can, even though its own deduction would be 'vainly sought', serve 'as the principle of the deduction of an inscrutable faculty which no experience could prove . . . namely the faculty of freedom' (*CPracR* 5: 47; Gregor, 178). But this is apparently to concede that the categorical imperative cannot be demonstrated by the methods of transcendental philosophy at all, even if once it is granted the reality of freedom might be derived from it by an argument that employs the method of transcendental philosophy; and this is a result on which Fichte attempts to improve.

Second, Kant supposes that we human beings are not just immediately aware of the binding force of the moral law in the form of the categorical imperative, but that we are aware of it by means of or at least in conjunction with a distinctive *feeling*, the feeling of 'respect' or 'reverence' (*Achtung*). He claims that 'though respect is a feeling, it is not one *received* by means of influence; it is, instead, a feeling *self-wrought* by means of a rational concept and therefore specifically different from all feelings of the first kind, which can be reduced to inclination or fear' (*G* 4: 402n.; Gregor, 56), and even that 'this feeling is the only one that we can cognize completely *a priori* and the necessity of which we can have insight into' (*CPracR*, 5: 73; Gregor, 200). But even if we were to overlook Kant's first

³ Immanuel Kant, *Groundwork of the Metaphysics of Morals*, 5: 413; cited from Kant, *Practical Philosophy*, tr. and ed. Mary J. Gregor (Cambridge: Cambridge University Press, 1996), 66. The *Groundwork* will henceforth be cited as *G* and the *Critique of Practical Reason*, also to be cited in Gregor's translation from *Practical Philosophy*, as *CPracR*.

problem and suppose that he does have a theory of our a priori knowledge of the content and the binding force of the moral law itself, it is not at all clear on what grounds he thinks that we can know a priori that this moral law must make its presence and force known *through a feeling*, that is, it is not clear how we know not just that this feeling 'has an intellectual cause, which is known a priori' (*CPracR* 5: 79; Gregor, 204), but how we also know a priori that this intellectual cause must produce a specific feeling. Fichte attempts to improve upon this aspect of Kant's foundations of morality also.

Finally, Kant has a problem about the compatibility of his proof of the reality of freedom with the undeniable fact that human beings do not always act as morality demands. At least he has such a problem in the *Groundwork*, where his proof that the moral law is the law of our noumenal selves because the noumenal self is essentially rational amounts to a proof that the moral law is the *causal* law of the noumenal self, a 'causality in accordance with immutable laws but of a special kind' (*G* 4: 446; Gregor, 94), which, if the noumenal self is the complete ground of the phenomenal self, as Kant at least initially supposes when he states that 'the world of understanding contains the ground of the world of sense and so too of its laws' (*G* 4: 453; Gregor, 100), would in turn mean that the phenomenal self also cannot but act in accordance with the moral law, that is, cannot act immorally, although obviously people do. As soon as he has introduced his transcendent proof that the moral law is the law of our noumenal will in *Groundwork*, III (*G* 4: 451–2; Gregor, 98–9), Kant attempts to avert this implication by writing as if the moral law of our noumenal selves is only an *ideal* for our phenomenal selves, and as if our actions in the phenomenal world are the outcome of a struggle between our noumenal selves governed by the moral law and our phenomenal selves that 'conform wholly to the natural law of desires and inclinations' (*G* 4: 453; Gregor, 100), or between our 'proper' or authentic and our inauthentic and lesser selves (*G* 4: 457; Gregor, 104). But obviously the treatment of our inclinations as a separate source of agency not grounded in our noumenal selves is inconsistent with the supposition that our noumenal selves are the complete ground of our phenomenal selves, and Kant's theory of freedom in the *Groundwork* is incoherent. This problem was quickly pointed out by several writers,⁴ and Kant attempted to resolve it by the time he wrote the *Religion within the Bounds of Mere Reason* (1793) by retaining the supposition that the noumenal self is the complete ground of the phenomenal self but jettisoning the supposition that the moral law is the *causal* law of the noumenal self, supposing instead that the noumenal self can freely choose (through its *Willkür* or faculty of choice) whether to subordinate the moral law (which is given to it by pure practical reason or *Wille*) to the principle of self-love or rather to prioritize self-love

⁴ The objection that Kant's identification of the moral law with the causal law of the (noumenal) will makes free but immoral action inexplicable, often thought to have been raised by Karl Leonhard Reinhold in 1792, was in fact first raised by Johann August Heinrich Ulrich in his *Eleutheriologie, or on Freedom and Necessity*, in 1790. Carl Christian Erhard Schmid and Reinhold then both attempted to respond to it with distinctions similar to Kant's own distinction between *Wille* as the source of the moral law and *Willkür* as the ability to choose whether or not to conform to that law, the former in his *Versuch einer Moralphilosophie* of 1790 and the latter in the Eighth of his *Briefe über die Kantische Philosophie*, which appeared in 1792 but only after Kant had published the first part of his own *Religion within the Boundaries of Mere Reason* in January of that year.

over the moral law, and that the phenomenal self will faithfully reflect this noumenal choice. This renders the possibility of freely choosing to do evil at least *conceivable* or *consistent* with Kant's theory of the noumenal basis of the phenomenal self, but at the cost of rendering the nature of noumenal choice, to use Kant's own term, *inscrutable*, and it certainly puts the last nail in the coffin of any attempt to derive the validity of the moral law from the essential character of our noumenal selves. Fichte will attempt to do better than Kant on this issue as well.

Determining whether Fichte does any better than Kant in applying the transcendental method to any foundational issue in moral philosophy depends, of course, on how we understand this method in the first place. What Kant meant by transcendental method has of course been long discussed and debated; what Fichte means by such a method has been more recently discussed.⁵ For purposes of the present discussion of Fichte's attempt to improve upon Kant's transcendental moral philosophy in the *System of Ethics* of 1798, it seems most natural to derive one's conception of his transcendental method in general not only from this work itself but also from the version of the *Wissenschaftslehre* that Fichte promulgated during the period of the composition of this work, namely that preserved in the lectures that he gave during the period 1796–9, now known as the *Wissenschaftslehre nova methodo*.⁶ Two points that are essential to the method of transcendental philosophy as Fichte conceives of it in the period represented by this work may be emphasized here. First, transcendental philosophy aims to describe the structures of thought that make self-consciousness possible, structures that will include awareness of the self's activity on representations thought of as originating from something without the self but where the latter thought is recognized to be part of the self's own thought: transcendental philosophy is thus supposed to be immanent rather than transcendent, that is, to confine itself to describing the structures of thought that are the conditions of the possibility of experience rather than anything outside of experience, even though experience itself includes the thought of something outside of itself. As Fichte puts it in the *Wissenschaftslehre nova methodo*:

We are not talking about anything more than the occurrence of this representation [of the self]! . . . Here we are dealing with nothing but an immediate positing of the I, and this is a representation. The idealist's principle is present within consciousness, and thus his philosophy can be called 'immanent' . . . Dogmatism is transcendent; it soars beyond consciousness. Idealism is *transcendental*; for though it remains within consciousness, it shows how it is possible to go beyond consciousness. That is to say, it shows how we come to assume that there are things outside of ourselves which correspond to our representations. (WNM 94–5)

In other words, transcendental philosophy concerns itself with the self's *representation* of its relation to objects other than itself, but does not concern itself directly with those

⁵ See esp. work by Daniel Breazeale, such as 'The "Mixed Method" of Fichte's *Grundlage des Naturrechts* and the Limits of Transcendental *Reellephilosophie*', in Tom Rockmore and Daniel Breazeale (eds), *Rights, Bodies and Recognition: New Essays on Fichte's Foundations of Natural Right* (Aldershot: Ashgate, 2006), 117–37, and ch. 4 in the present volume.

⁶ Fichte, *Foundations of Transcendental Philosophy: (Wissenschaftslehre) nova methodo* (1796/99), tr. and ed. Daniel Breazeale (Ithaca, NY, and London: Cornell University Press, 1992); cited as WNM.

things apart from the fact of their being represented by the self. Or as he puts in *The System of Ethics* itself, 'Anyone familiar with the spirit of transcendental philosophy will share our presupposition that this thinking of something subsisting must itself be based on our laws of thinking and that, accordingly, what we are seeking is only the essence of the I for the I, and by no means the latter's essence in itself, as a *thing* in itself' (SE 33). In Fichte's language, 'the essence of the I for the I' refers to the conditions for the possibility of self-consciousness, and in this statement Fichte is essentially endorsing Kant's own bipartite conception of transcendental philosophy as consisting in the analysis of the conditions of the possibility of self-consciousness on the one hand (the core of the Transcendental Analytic of the *Critique of Pure Reason*) and the critique and rejection of the transcendent metaphysics of the substantial self on the other (the topic of the Paralogisms of Pure Reason).

Second—and here is where Fichte departs from and aims to improve upon Kant's conception of transcendental philosophy—what transcendental philosophy reveals is that the essence of the self's thought of both itself and the things beyond itself that it represents is its *activity*, that representing and thinking are forms of activity, and that this cannot be understood except by thinking of the self as an agent in the world, or from the practical standpoint. 'The I is nothing but its own activity,' Fichte states, 'The representing subject is identical with its own self-activity, which constitutes its very essence, and thus, in every specific situation, its essence consists in a certain, specific self-activity' (WNM 97). But particularly in the *Wissenschaftslehre nova methodo*, Fichte stresses that concrete human individuals can only come to understand the active nature of thought through understanding what is distinctive in action in the ordinary sense, goal-directed bodily intervention in the course of the world, so that 'Considered from an a priori perspective and within the context of its place within a genetic account, the [philosophical] viewpoint is found to arise in the course of acting, and thus it can also be called "the practical point of view"' (WNM 106). The key to understanding the nature of self-consciousness in general thus becomes the understanding of human action, and the key to understanding this is understanding freedom. The key to understanding freedom, in turn, is to understand that activity must have its own law distinct from the laws that govern that which is represented merely as object, and the key to Fichte's transcendental derivation of the moral law is then the insight that the moral law is the only candidate for such a law of the distinctive activity of the self: the key to Fichte's system of ethics but also to his entire philosophy in the period 1796 to 1798 or 1799 is thus the argument that understanding the self requires understanding its activity but understanding its activity requires understanding its subjection to the moral law. For this reason, Fichte emphasizes in this period that he no longer recognizes the traditional distinction between theoretical and practical philosophy. The first presentation of his philosophy in 1794, Fichte now writes,

was made somewhat awkward by the fact that the discussion of the conditions for the possibility of the principles [of transcendental philosophy] did not present these conditions in their

natural order, but was instead divided into a 'theoretical' and a 'practical' part. As a result of this division, many directly related issues were separated too widely from one another. This will no longer occur in the present version, [which will follow] {a method of presentation that is just the opposite of that followed by the author in his compendium of 1794, where he proceeded from the theoretical portion of philosophy (i.e., from what had to be explained) to the practical part (i.e., to what was meant to serve *as the basis* for explaining the former). In the present lectures, however, the hitherto familiar division between theoretical and practical philosophy is not to be found. Instead, these lectures present philosophy *as a whole*, in the exposition of which theoretical and practical philosophy are united. This presentation follows a much more natural path, beginning with the practical sphere . . . in order to explain the [theoretical] in terms of the former . . .} (WNM 85–6)

If *The System of Ethics* is truly composed in the spirit of this methodological remark from the *Wissenschaftslehre nova methodo*, then it should not be considered as an addendum to Fichte's transcendental philosophy but as its foundation, and his attempt to provide a satisfactory transcendental derivation of the central concepts of practical philosophy should not be regarded just as an attempt to improve upon Kant's use of the transcendental method in practical philosophy but as an attempt to improve upon Kant's use of the transcendental method *überhaupt*.

2. The Transcendental Deduction of the Categorical Imperative

This is not the place for a detailed discussion of the general argument of the *Wissenschaftslehre nova methodo*; I will comment only briefly on the structure of Fichte's general argument that awareness of activity is the condition of the possibility of the awareness of selfhood and then discuss in more detail his argument in *The System of Ethics* that recognition of the moral law is the condition of the possibility of awareness of activity. The overall argument is stated compactly in a series of 'dictata' with which Fichte summarized his lectures during the course of giving them and which the transcriber of the lectures, Karl Christian Friedrich Krause, placed at the beginning of his transcription of them (WNM 35, 65).

Here Fichte states that the philosophical task at hand is to 'Construct the *concept of the I* and observe how you accomplish this.' Fichte then claims that if one does 'one will discover that one is *active* and will discover in addition that one's activity is directed upon *one's own active self*. Accordingly, the concept of the I comes into being only by means of a *self-reverting activity*' in which '*one posits oneself as self-positing*' (WNM 65). By this, Fichte means that representing an object is a form of activity and representing oneself as representing an object is also a form of activity: thus understanding oneself as a self representing a world of objects requires understanding the activity of representing objects and the activity of representing oneself as representing objects. Fichte stresses that the mental activity that is involved twice over here cannot

be reduced to a passive event of copying or merely receiving an impression; rather, 'The activity involved in this transition is called *real activity* and is opposed to that *ideal activity* which merely copies the former' (WNM 67). Fichte does not pause at this stage in his summaries to spell out in exactly what way or ways the self must be active in order to represent objects or to represent itself representing objects, although he subsequently illustrates at least part of why representing is active, tacitly drawing on Kant's Transcendental Aesthetic, by arguing that the first condition for the possibility of representing objects is representing them as occupying different places in space and that this in turn requires the activities of *constructing* ('drawing') space itself and representing or *placing* objects at different locations in the space thus constructed. He does not, however, directly allude to the argument of Kant's Transcendental Analytic that *judging* that different representations are representations of a single object also requires the *activity* of combining or synthesizing representations, so that since knowledge requires judgement there is no knowledge without activity. Rather, what he does do is stress that any form of activity at all, of which self-consciousness has been asserted to involve at least two instances, requires a conception of oneself as *free*, as *self-determining*, as oneself determining how one's mental capacities are to be exercised rather than as merely being acted upon by external agencies. His claim is that 'Freedom is therefore the ultimate ground and the first condition of all being and of all consciousness' (WNM 68). The starting-point of Fichte's transcendental philosophy in the period of the *Wissenschaftslehre nova methodo* and *The System of Ethics* is thus the claims that understanding oneself as a self involves understanding oneself as not merely having representations but as acting upon representations, and that this in turn requires understanding oneself as free or self-determining.

Fichte's next, and for our purposes, crucial claim is that understanding oneself as self-determining requires the concepts of practical philosophy and ultimately the recognition of oneself as governed by the moral law rather than by mere laws of nature. The first of these steps is asserted in the *Wissenschaftslehre nova methodo* and the latter is spelled out in *The System of Ethics*. In the former work, Fichte states that 'Free self-determination is intuitable only as a determination to become "something", of which the self-determining or practical {power} must possess a {freely constructed} concept', where in turn 'A concept of this sort is called "the *concept of a goal*". Consequently, for the intuiting subject, the same subject who possesses *practical power* at the same time possesses the power to form concepts, just as, conversely, the *comprehending subject*, or {the power of} *the intellect*, must necessarily be practical' (WNM 68). Fichte's claim here is that the activity of representing can only be understood as the act of determining one's general capacity or potential—determinability—to represent in some specific or determinate way, to represent some specific object, and that this must in turn be conceived of as realizing a goal or engaging in a goal-directed activity. 'A *free action* is possible only if it is guided by a freely constructed concept of this action', or a freely chosen goal. The next task is then to understand what it takes to

represent oneself as a being that can choose a goal and choose to act in a way intended to realize such a goal.

Fichte develops the second stage of this argument, that one can represent oneself as freely choosing a goal only by representing oneself as subject to the moral law, the argument by which he intends to resolve the first challenge in improving upon Kant's use of the transcendental method in practical philosophy, in the Introduction to *The System of Ethics*. The Introduction begins with a statement of the 'direction of fit' contrast that Elizabeth Anscombe was to make well-known among Anglo-American philosophers a century and a half later: '*The entire mechanism of consciousness rests on the various aspects of [the] separation of what is subjective from what is objective, and, in turn, on the unification of the two*,' he begins; there are two ways in which the subjective and the objective may be unified, he continues, namely in cognition, where the subjective is supposed to agree with the objective, and 'when *I act efficaciously (wirke)*', where 'the two are viewed as harmonizing in such a way that what is objective is supposed to follow from what is subjective', where 'a being is supposed to result from my concept (the concept of an end)' (SE 7–8). But in fact the activity of the self must be recognized in order for the distinction between subjective and objective even to be made in the first place, for self can be distinguished from object only by recognition of one's activity in the world: 'I find myself to be acting efficaciously in the world of sense. All consciousness arises from this discovery. Without this consciousness of my own efficacy, there is no self-consciousness; without self-consciousness, there is no consciousness of something else that is not supposed to be I myself' (SE 8–9). Presumably Fichte's thought is that if I were entirely passive, just a mirror for changes going on in an entirely objective world, I would have no way of distinguishing myself from that world: a mirror, after all, which passively receives images, is just another piece of furniture of the world. It is only because I can initiate changes in the world, myself determine what happens there in accordance with my own concepts of my goals, that I can have a sense of myself as distinct from the rest of the objects of the world in the first place, and then in turn conceive of objects as distinct from that self. In order to have a sense of myself as distinct from the world, I have to have not only 'a representation of the *stuff*' of the world but also a representation of the '*properties* of this stuff . . . that are changed by my efficacy' and 'a representation of this *progressive process of change*, which continues until the shape that I intend is there' (SE 9). Thus, Fichte begins with the claim that the most fundamental feature of self-consciousness, the very distinction between subjective and objective, presupposes a conception of the purposive activity of the self, because it is only by conceiving of itself as acting on the world that the self can distinguish itself from the world. 'Accordingly, insofar as I know anything at all I know that I am active.'

The key move in Fichte's deduction of the moral law is then the argument that being conscious of oneself as active requires being conscious of oneself as acting not merely in accordance with a concept of an end or purpose, but with a concept of an end or purpose that is independent of and self-sufficient from any merely objective laws of nature, because if one did not so conceive of one's activity, one's conception of it would

collapse back into a representation of oneself as a merely passive object of nature, which would not be sufficient for a conception of one's own consciousness after all. Fichte claims that 'the principle of all practical philosophy' is derived from the explanation of 'how I come to assume that something objective follows from something subjective, a being from a concept'. 'This assumption arises', he states, 'because I have to posit myself absolutely as active, and, since I have distinguished something subjective and something objective within myself, I cannot describe this activity otherwise than as the causality of a concept . . . Absolute activity in this shape is also called *freedom*. Freedom is the sensible representation of self-activity, and it arises through opposition to the constrained state both of the object and of ourselves as intelligence, insofar as we relate an object to ourselves' (SE 14). If I were to try to conceive of myself as acting in according with concepts of ends but were to suppose that those concepts of ends were simply given to me from elsewhere, I could not conceive of myself as anything but another part of the objective world, and thus could not conceive of myself as active and as a self after all; I can only conceive of myself as a self acting on the world insofar as 'I presuppose a concept designed [or "projected": *entworfen*] by myself'. Fichte continues, 'the concept of an end, as it is called, is not itself determined in turn by something objective but is determined absolutely by itself. Were this not the case, then I would not be absolutely active and would not be immediately posited in this way; instead, my activity would depend on some being and be mediated by that being—which contradicts our presupposition', that is, the presupposition that I am active, which is in turn the presupposition of my recognition of myself as a self at all. Fichte concludes:

The most important result of all this is the following: *there is an absolute independence and self-sufficiency of the mere concept* (that which is 'categorical' in the so-called categorical imperative), due to a causality of what is subjective exercised upon what is objective—just as there is supposed to be an absolutely self-positing *being* (of the material stuff), due to a causality of what is objective exercised on what is subjective. With this we have joined together the two extremes of the entire world of reason. (SE 15)

In order to conceive of ourselves as acting, we have to conceive of a world of stuff on which to act, things with properties that we have to modify, and no doubt we have to conceive of that stuff as having its own laws with which we have to cope; but we cannot conceive of ourselves as also governed entirely by those laws, because then we will not be able to conceive of ourselves as different from and acting on that stuff after all. We must conceive of our action as governed by the principle of acting in accordance with or for the sake of ends that are determined independently of the mere laws of stuff, thus for the sake of ends that are 'independent' from and 'self-sufficient' with regard to the mere laws of stuff.

This is the gist of Fichte's deduction of the moral law, but he amplifies his presentation of it in Part I, the explicit 'Deduction of the Principle of Morality'. Here Fichte presents his premise that I must think of myself as acting in order to think of myself as a self at all as the 'Theorem' that 'I FIND MYSELF AS MYSELF ONLY AS WILLING' (SE 24), and his

deduction then proceeds by means of a description of the conditions of the possibility of willing; his argument is thus analogous to Kant's derivation of the categorical imperative from the concept of a rational will in section II of the *Groundwork*, with the key difference that whereas Kant presents his analysis as yielding merely analytical truths about any possible rational will that can be shown to apply to us, thus to be synthetic a priori, only if he can give a further proof that we actually are rational beings with wills, Fichte has already paved the way for our acceptance of the results of this analysis by means of his previous argument that the recognition of our activity is the condition of the possibility of self-consciousness itself. Fichte's analysis of willing proceeds thus. On the one hand, 'WILLING ITSELF . . . IS THINKABLE ONLY UNDER THE PRESUPPOSITION OF SOMETHING DIFFERENT FROM THE I', namely, something to be acted upon but also some particular way in which that external thing is to be acted upon, because 'All willing that is actually *perceivable* . . . is necessarily a determinate willing, in which *something* is willed.' Thus, 'what I will is never anything but a modification of an object that is actually supposed to exist outside of me' (SE 29). On the other hand, 'IN ORDER TO FIND MY TRUE ESSENCE I MUST . . . THINK AWAY ALL THAT IS FOREIGN IN WILLING. WHAT THEN REMAINS IS MY PURE BEING' (SE 30); in other words, I cannot think of the determinate change that I will in an external object as dictated by that object, because then I will not be thinking of myself as willing after all. Thus, 'Insofar as willing is something absolute and primary, . . . it simply cannot be explained on the basis of any influence of some thing outside the I, but only on the basis of the I itself; and *this absoluteness* of the I is what would remain following abstraction from everything foreign' (SE 30). So 'THE ESSENTIAL CHARACTER OF THE I, THROUGH WHICH IT DISTINGUISHES ITSELF FROM EVERYTHING OUTSIDE OF IT, CONSISTS IN A TENDENCY TO SELF-ACTIVITY FOR SELF-ACTIVITY'S SAKE' (SE 34); in other words, in order to think of the self as willing, it must have an end, but in order to think of it as genuinely active, this end cannot be imposed upon it by external objects and their laws, so the only alternative is to think of the self itself as an end—in other words, to think of the self as an end in itself, and of that conception as the ground of the law in accordance with which the self acts.

This argument is clearly supposed to parallel Kant's argument that rational being as an end in itself is the ground of a possible categorical imperative, but to derive that rational being is an end in itself from the conditions of the possibility of self-consciousness rather than to merely assert it, as Kant does (G 4: 428; Gregor, 78–9). But it might seem open to the objection that it is only *my own* self-activity that it makes into the non-externally-imposed end of my willing, and thus that it does not give rise to a genuine moral law, equivalent to Kant's second formulation of the categorical imperative as 'So act that you use *humanity*, whether in your own person or in the person of any other, always at the same time as an end, never merely a means' (G 4: 429; Gregor, 80), after all. I take it that Fichte's assumption is that, in arguing that self-consciousness must distinguish the end of its willing and action from any particular end merely imposed upon it by particular external, empirical circumstances, he has already shown that a genuine will must will a general rather than merely personal end, thus

that self-activity in general, whether in one's own empirical person or that of any other, must be the ultimate end of willing, although of course an end that can lead to determinate action only insofar as it implies some determinate object of willing in one circumstance or another. I think that this thought is implicit in Fichte's characterization of the results of his argument as applying to 'a rational being,' that is, to *any* rational being. Thus he writes, 'Strictly speaking, our deduction is now concluded. Its proper and final goal was, as we know, to derive from the system of reason as such the necessity of thinking that we ought to act in a certain manner and to demonstrate that if any rational being whatsoever is assumed, such a being must think such a thought,' and he continues:

A rational being is *itself* supposed to produce everything that it is ever actually to be. You therefore have to ascribe to such a being some sort of existence prior to all actual (objective) being and subsistence . . . This manner of existing can be none other than existing as an intellect in and with concepts. In your present concept [of yourself] you therefore must have thought of a rational being as an intellect . . . (SE 52–3)

There is nothing personal in the thought of oneself as a rational being, although that thought is a condition of the possibility of self-consciousness; therefore there is presumably nothing personal in the thought of self-activity as the ultimate end of willing one's end as a rational rather than merely natural being, either. I think that Fichte makes the same point more explicitly when he argues that the principle that is necessary in order to conceive of oneself as willing, thus as active, thus as self-conscious, and that is imposed upon oneself to this end, cannot be conceived of as a self-referential rule suggested by particular circumstances, because then it would not be a rule of genuine activity after all, but can only be a general rule furnished by the pure intellect. Thus, beginning his somewhat protracted conclusion of Part I of *The System* (like a Bruckner symphony, the section seems to reach repeated climaxes before it finally ends), he states that 'the thought just deduced has been called a "*law*" or a "*categorical imperative*"; and proceeds to explain why this is appropriate:

We are able to think of freedom as standing under absolutely no law, but as containing the ground of its determinacy purely and entirely within itself—the determinacy of a thinking that is subsequently thought of as the ground of a being; and this is how we must think freedom if we want to think it correctly, for its essence lies in its concept, and the latter is absolutely indeterminable through anything outside itself. Since what we are thinking of is freedom, and freedom is determinable in all possible ways, we can also think it as subject to a hard and fast rule. The concept of such a rule, however, is something that only a free intellect could design for itself, and only a free intellect could freely determine itself in accordance with such a rule. The intellect could thus make for itself a great variety of different rules or maxims—for example, rules pertaining to self-interest, laziness, the oppression of others, and other similar rules—and could obey these rules steadfastly and without exception, and always freely. Let us now assume, however, that the concept of such a rule imposes itself on the intellect, i.e., that the intellect is, under a certain condition, required to think a certain rule, and only this rule, to be the rule governing its own determinations. We may rightly assume something of this sort, since the intellect, thought

absolutely free with regard to the sheer occurrence of an act of thinking, still stands under determinate laws with regard to its way and manner of thinking.

In this way, the intellect would be able to think of a certain way of acting as conformable to the rule and another way of acting as contradicting it. (*SE* 56–7)

Here Fichte seems to be arguing as follows. It might seem possible to think of such self-referential maxims as to serve self-interest, to indulge in laziness, or to oppress others when one wants—examples that seem quite carefully chosen to echo the maxims that are to be excluded by Kant's own illustrations of the categorical imperative, specifically his second through fourth illustrations (see *G* 4: 422–4 and 429–31; Gregor, 74–5 and 80–1)—as freely chosen rules. But in fact they could not furnish genuine rules that could be genuinely observed or violated, because since they are merely self-referential and arbitrary, they can be changed at any time. Instead, what is to serve as a genuine rule for the free intellect and will must be something that such an intellect and will chooses to follow independently of the mere desires that would ground such self-referential maxims, and such a rule could only be genuine, valid for any free intellect and will, and comprehending all free intellects and wills in its scope. 'The concept of such a rule is something purely and simply primary, something unconditioned, which possesses no ground outside itself, but is grounded completely in itself. Hence the action in question is not one that ought to occur for this or that reason, or because something else has been willed . . . instead, this is an action that ought to occur purely and simply because it ought to occur' (*SE* 57–8). Genuine freedom can be obtained only through generality, thus if freedom is a condition of the possibility of self-consciousness then so is generality.

In addition to explaining how his own deduction of the moral principle satisfies Kant's demand for a categorical imperative, he also tries to show how his argument yields a principle of autonomy (*SE* 58), thus further confirming his own ambition to give a transcendental argument for the moral law as Kant conceives it. I will not comment on this further step, however, but will instead conclude this section with a general comment on Fichte's strategy for a transcendental method in practical philosophy. Kant, as I earlier noted, seemed to recognize only two options with regard to a proof of the moral law: either it needs to be grounded not by a transcendental proof but by a step into transcendent metaphysics, that is, the argument of *Groundwork*, III, that the moral law is equivalent to the law of reason and is the law of our noumenal selves because the noumenal self is essentially rational, or else the moral law cannot be grounded at all, but can only be accepted on faith, as a 'fact of reason', although once accepted it can ground an argument for the reality of freedom. Fichte, however, has attempted to argue that the recognition of self-activity as the end of willing for its own sake is necessary in order properly to conceive of oneself as willing, and that to conceive of oneself as willing is a condition of the possibility of self-consciousness itself. If an argument to the condition or conditions of the possibility of self-consciousness itself is the paradigmatic form of a transcendental

argument, as could well be argued if Kant's own transcendental deduction of the pure concepts of the understanding is supposed to be the paradigm of a transcendental argument, then Fichte's deduction of the moral law does have the form of a transcendental argument for the moral law, which eluded Kant. I have not, of course, attempted to demonstrate that Fichte's transcendental argument for the moral law is compelling, although I have tried to show how Fichte can be seen as having attempted to avert one obvious objection to it, namely the objection that it merely makes the recognition of one's own self-activity as an end in itself the condition of the possibility of one's own self-consciousness. Fichte can be thought to have argued, I have suggested, that it is only by conceiving of oneself as willing an entirely general end that one can conceive of oneself as genuinely willing at all, and not merely as being acted upon by contingent and idiosyncratic desires. The key to the success of this defence is his underlying assumption that one needs to conceive of oneself as genuinely acting in order to conceive of oneself as a self at all, and that to conceive of oneself as being moved by merely accidental desires would not be to conceive of oneself as genuinely acting at all.

This is, to be sure, far from a complete assessment of the success of Fichte's attempt to provide a genuinely transcendental argument for the moral law. Instead of attempting to assess his argument, however, I now want to turn to the second aspect of his attempt to improve upon Kant, namely his attempt to give a transcendental argument for the existence of moral feeling.

3. The Transcendental Argument for Moral Feeling

Kant, as I noted, claimed that the feeling of respect is 'self-wrought' by reason and the only feeling that we can in some way know a priori, but he certainly did not explain the latter claim. A second distinctive feature of Fichte's attempt to apply the transcendental method to practical philosophy is that he does attempt to give an argument that the moral determination of the will must be accompanied with a distinctive feeling, indeed that the moral determination of the will must make itself effective *through* a distinctive kind of feeling and the relation of that feeling to other feelings. And while Kant, for example in the *Critique of Practical Reason*, treated the feeling of respect as a pendant to the moral law, an 'incentive' that could be described after the presentation of the 'fact of reason' or consciousness of the moral law and the deduction of freedom from it had been completed, Fichte includes the premise for his account of moral feeling in his original deduction of moral law and his fuller account of it under the rubric of a 'Deduction of the Reality and Applicability of the Principle of Morality' (SE 65): his thesis is that human freedom and its direction by the concept of the end of activity for its own sake can only be empirically manifest to the human being in the form of a distinctive feeling, a feeling that stands in a distinctive relation to other feelings, so that the existence of a distinctive form of moral feeling is not an unexplained accident but

a condition of the possibility of consciousness of the self as a moral being and agent at all. In this sense Fichte attempts to give a transcendental deduction of moral feeling.

There are two underlying premises of Fichte's argument. The first is that the moments of self-consciousness, including both what can be thought of as its content and as the activity it performs on that content, must be manifest to consciousness in some empirical form, that is to say, in some spatio-temporal form. The second is that self-consciousness requires the unification of the objective and the subjective, but a unification that is really a reunification, that is, a unification of something objective and something subjective that are really already within the self. Putting these together, he reaches the conclusion that both the objective and the subjective moments of consciousness must be represented to us in some temporal form. In the context in which it has been customary for us to start thinking about self-consciousness in the philosophical tradition, namely the context of the perception of external objects, we think of the objective as presented by sensations and of the subjective as manifested in the activity of organizing and conceptualizing these sensations, and here it would seem that Fichte's general desideratum could be satisfied by supposing that we have some form of empirical awareness of our activity of conceptual organization as well as of sensation—the former being precisely what he called, so misleadingly from a Kantian point of view, 'intellectual intuition', by which he means not an intuition of objects of pure reason but an intuition of the activity of reasoning. In the context that Fichte himself thinks is primary for the explication of self-consciousness, namely, the consciousness of our own activity, there must also be two forms of empirical consciousness, namely some empirical consciousness of freedom itself, but also some consciousness of that on which freedom is exercised—the objective for freedom as subjective, that which is analogous to the sensation that is organized by conceptual activity in the cognitive case. Fichte first introduces the empirical manifestation of freedom itself as the subjective moment in practical activity. He maintains that 'THE I BECOMES CONSCIOUS OF ITS OWN TENDENCY TOWARD SELF-ACTIVITY . . . IN RELATION TO THE ENTIRE I AS A DRIVE' (SE 43–4). A drive is a temporally extended and forward-looking state that is the empirical manifestation of freedom. But although self-activity or freedom is the essence of the self, it is also only part of the self, and so the drive is only a tendency, something that is potential as well as actual, not fully realized, and therefore the empirical manifestation of freedom must also in some way manifest the incompleteness of freedom at any given time. In his first introduction of the existence of the drive that manifests self-activity in the original 'Deduction of the Principle of Morality' itself, Fichte uses this point only to infer that 'the drive will be accompanied by a longing' that will be resolved at least to some extent when a 'deed is accompanied by a decision' (SE 44). But in the fuller discussion in the 'Deduction of the Reality and Applicability of the Moral Principle' Fichte clarifies that, since activity requires something that it transforms, freedom must overcome some resistance, and then since both freedom and the resistance that it has to overcome must be empirically manifested in feelings, there must be feelings that have to be resisted as well as the feeling of

the longing for the freedom to resist and overcome then. For this reason moral action must take the form of the attempt of moral feeling to overcome or redirect other feelings. Fichte makes these points in the crucial 'Proof' and 'Corollaries' of this section. First, the key claims of the proof:

I find myself to be willing only insofar as my activity is supposed to be set in motion by a determinate concept of the same . . .

Activity cannot be determined by itself, and yet it must be determined if consciousness is to be possible at all. This means nothing else than the following: the activity is to be determined through and by means of its *opposite* . . .

I cannot however, intellectually intuit, absolutely and by myself, the manner in which I am limited; instead, this is something I only *feel* in sensory experience . . . The I is now supposed to be posited as active; and thus it would have to be posited as eliminating and breaking through a manifold of boundaries and resistance, in a succession . . . (SE 88–9)

This argues that activity is made determinate only by overcoming some specific form of resistance, and that I must feel in sensory experience the specific resistance or limits that I have to overcome in my activity as well as feeling, as already argued, the tendency of my activity to overcome them. Fichte reiterates this point in his 'Corollaries':

The I is to be posited as an actual I, but solely in contrast with or in opposition to a Not-I. But there is a Not-I for the I only under the condition that the I acts efficaciously and feels resistance in its effective operation, which, however, is overcome, since otherwise the I would not be acting efficaciously. Only by means of such resistance does the activity of the I become something that can be sensed and that endures over a period of time, since without such resistance the I's activity would be outside of time, which is something we are not even able to think. (SE 89)

Here Fichte argues that the essential nature of the self and the basis for self-consciousness is not merely the *positing* of a Not-I by the I, as he had in his original presentation of the *Wissenschaftslehre* in 1794, but the *overcoming* of the Not-I by the I, and that both moments, the Not-I and its overcoming, must be *felt* because self-consciousness takes place in time and feeling is the form of the presentation of the moments of self-consciousness in time. But it also should be clear from everything that has preceded the argument that what he is here calling 'Not-I' is not something that is simply outside the self, but rather that *in the self* which seems to resist what is essential in the self, its free activity: so the 'Not-I' here can only stand for *feelings* that resist the tendency and longing to morality but that can be overcome or redirected under the guidance of the moral law. This leads to an empirical model of free, moral action in which feelings resistant to morality must be controlled by the longing to be free and moral in general, which Fichte further amplifies by suggesting that specific feelings resistant to morality must be replaced by others compatible to or hospitable with morality:

What ordinary consciousness tells us therefore is this: in executing our ends, we are bound to a certain order of means. What does this claim mean when one views it from the transcendental viewpoint and when one attends solely to the immanent changes and appearances within the

I, in total abstraction from things outside of us?—According to the preliminary elucidations provided above, whenever I perceive I *feel*. ‘I perceive changes outside of me’ means that the state of my feelings has changed within me. ‘I want to act efficaciously outside of myself’ means that I will that the place of one determinate feeling should be occupied by another determinate feeling, which I demand through my concept of an end. ‘I have become a cause’ means that the feeling that is demanded actually does occur. Thus, ‘I proceed to my end by passing through the means’ means that other feelings occur in the interval between the feeling from which I proceed to willing and the feeling demanded by my willing . . .

Every feeling, however, is an expression of my limitation; and ‘I possess causality’ always means that I expand my limits. Thus what we are claiming is that this expansion can occur only in a certain, progressive series, since we are claiming that our causality is limited to the employment of certain means for accomplishing our end . . . (SE 92–3)

Here Fichte suggests that acting is always a temporally extended process, and thus that the act of overcoming resistance that presents itself in the form of feelings must take the form of successively replacing those feelings with others more conducive to the ultimate goal of the action.

The steps in this extended argument are no doubt loose, but its ambitions are interesting. Kant assumed that moral choice, for him the determination of the noumenal will, must have consequences in the empirical world, and he devoted considerable effort in the second and third critiques to explaining the conditions of the possibility of the realization of the ultimate objective of moral choice, namely the highest good, in the natural world. But his account of the mechanisms of human psychology by means of which the moral determination of the will is translated into action in the spatio-temporal realm of nature is sketchy and unargued. In the *Groundwork*, he tended to treat the feeling of respect as an epiphenomenal manifestation of the noumenal determination of the will to be moved by the moral law alone that plays no direct causal role in action (see G 4: 401n.; Gregor, 56). In the *Critique of Practical Reason*, he suggests that the feeling of respect which is itself caused by the moral law and supposedly known a priori to be so does play a causal role in action by striking down feelings of self-conceit (CPracR 5: 73; Gregor, 199), but he does not offer any argument that this role of the feeling of respect is a *sine qua non* of morally motivated action. In the *Metaphysics of Morals*, published as Fichte was working on *The System of Ethics* and not explicitly referred to in Fichte’s work, Kant does argue that the presence of a general moral feeling, conscience, and the more specific feelings of self-esteem and love of one’s neighbour are ‘antecedent predispositions on the side of *feeling*’ that are ‘conditions of receptiveness to the concept of duty’ without which we could not be ‘put under obligation,’ and that although we could not have a duty to somehow get these predispositions if we did not have them at all because in that case the concept of duty could get no purchase on us, we do have a duty to cultivate and strengthen them so that they will be forceful enough to be efficacious when we need them (*Metaphysics of Morals*, Doctrine of Virtue, Introduction, section XII, 6: 399–402; Gregor, 528–31). In a subsequent discussion of the specific feelings of ‘*sympathetic joy*’ and ‘*sadness*’ or ‘*sympathetic feeling*,’ Kant also says that such

feelings are the 'means' that nature has afforded us to accomplish our moral goals and that we have a 'particular, although only a conditional, duty' to 'cultivate the compassionate natural . . . feelings in us, and to make use of them as so many means to sympathy based on moral principles' (*Metaphysics of Morals*, Doctrine of Virtue, §§34–5, 6: 456–7; Gregor, 575)—a duty that is conditional, of course, because our duty is not simply to cultivate feelings of sympathy or other moral feelings and then to act on them whenever they happen to present themselves in sufficient strength, but rather to cultivate them so that they will be strong enough for us to act on in those circumstances but only those circumstances where the principle of morality tells us that the actions those then feelings prompt us to are indeed morally appropriate or mandatory.⁷ But Kant does not attempt to embed these eminently reasonable thoughts in any larger theory of human action, let alone to provide anything that might count like a transcendental argument for or deduction of them. That is exactly what Fichte has attempted to do by means of his extended argument (1) that self-consciousness requires a consciousness of free activity, (2) that consciousness of free activity requires a consciousness of both a goal to change reality and the resistance of reality, but especially one's own reality, to that change, (3) that all consciousness requires an empirical manifestation of its essential moments, and (4) that we must therefore have feelings that need to be overcome as well as a feeling of our drive or tendency to overcome them, and (5) that finally since action always takes place in temporally extended succession, (6) the act of overcoming feelings that are resistant to morality because of our general feeling or longing for self-activity—morality—must take the form of gradually substituting some specific feelings—which (7) might be analogous to Kant's specific moral feelings of self-esteem, love of one's neighbours, sympathetic joy and sadness, and so on—for non-moral or morally refractory feelings—the nature of which is apparently so obvious to all human beings that neither Kant nor Fichte feels any compulsion (pardon the pun) to enumerate them.

There are two assumptions in this argument that might be contested by someone pressing the question of whether this is really a transcendental argument. The first would be whether Fichte has genuinely shown the manifestation of moments of self-consciousness in empirical form to be a necessary condition of the possibility of self-consciousness itself, assuming that the strategy of transcendental argumentation is always to argue that its target is a condition of the possibility of self-consciousness itself. His underlying assumption is that we are creatures who are embodied in space and time (see, for example, *WNM* 70–1), and that all moments of self-consciousness must present themselves in some spatio-temporal form, indeed some form linked to our possession of bodies. It is not clear that he has given an explicit argument for this assumption. It is also not clear that we can seriously imagine any alternative to this assumption—even 'brains in the vat' arguments assume that we are at least body-parts,

⁷ For further discussion of the role of moral feeling in Kant, see my *Knowledge, Reason, and Taste: Kant's Response to Hume* (Princeton: Princeton University Press, 2008), ch. 4, and 'The Obligation to be Virtuous: Kant's Conception of the Tugendverpflichtung', *Social Philosophy and Policy*, 27 (2010), 206–32.

namely brains, located in some other bodies, namely vats, and only quibble about whether we can be sure about what sorts of bodies we really have and what surroundings they are really in—and thus not clear that any transcendental argument should really be under the burden of proof of showing that we have bodies at all. Perhaps any plausible transcendental argument has to start from the assumption that the conditions of the possibility of self-consciousness must have empirical and therefore bodily manifestations, and the real burden of proof that transcendental arguments should satisfy is only that of showing that specific conditions of the possibility of self-consciousness require specific sorts of bodily functions or in Kant's term 'predispositions'.

The second objection that could be raised about Fichte's argument would be whether he has really shown that something as specific as *feelings* is a condition of the possibility of the action that is in turn the condition of the possibility of self-consciousness itself. But here one might respond that, if feeling is taken in a sufficiently general sense, as a mental state that has some conceptualizable content but also involves some element of bodily awareness, then of course every moment of self-consciousness must for bodily creatures like us be manifested as some form of feeling. The question would then be not whether Fichte has succeeded in showing that self-consciousness depends upon feelings in general, but whether he has succeeded in showing that our self-consciousness as moral agents, which is supposed to be the necessary condition of self-consciousness in general, depends specifically on the feelings of a drive to be moral as well as feelings that resist that drive but can at least gradually be overcome by it. Here the first of Fichte's assumptions might seem more contestable than the second: that we have feelings resistant to the demands of morality ultimately seems too obvious to need to be proved, and what is interesting is primarily that Fichte has attempted to provide a transcendental proof of the fact that overcoming these feelings is always a temporally extended and never completed process; what is less obvious is that Fichte has succeeded in showing that there is a single moral feeling or 'longing' which is the start of that process. Nevertheless, it might still be concluded that if certain reasonable assumptions about the burden of proof in transcendental arguments are made, Fichte has shown how some key assumptions of moral psychology might be placed within the framework of a general analysis of the conditions of human action in a way that begins to look like a transcendental deduction of them.

4. The Problem of Freedom

Fichte seems to have made some advances over Kant in the use of a transcendental method for the deduction of the validity of the moral law and of moral feeling. It remains to ask whether he has made a genuine advance on Kant's treatment of the most fundamental problem in his theory of freedom, the apparent contradiction between the claim that the moral law is the causal law of the authentic self and

the indisputable reality of immorality. The course of Kant's thought on this problem about freedom from the *Groundwork*, his first work on practical philosophy, to the *Metaphysics of Morals*, his last, is a torturous one: to summarize it briefly, in the *Groundwork* he makes the moral law the causal law of the free noumenal self, making it impossible even to conceive how the free agent can choose to violate the moral law and do something immoral; in *Religion within the Boundaries of Mere Reason*, Kant introduces the vital distinction between *Wille* and *Willkür*, thus separating the will into one faculty that legislates the moral law and another that chooses whether or not to prioritize it over self-love, thus allowing for the possibility of immoral action even though any particular choice to act immorally remains inexplicable; but in the *Metaphysics of Morals*, Kant seems to remain uncomfortable with this resolution of the problem of the possibility of evil, arguing that 'freedom of choice cannot be defined—as some have tried to define it—as the ability to make a choice for or against the law (*libertas indifferentiae*), even though choice as a *phenomenon* provides frequent examples of this in experience' (*Metaphysics of Morals*, Introduction, section III, 6: 226; Gregor, 380), thereby appearing to take back precisely what he himself had argued in the *Religion*.⁸ It would hardly be surprising that Fichte should attempt to develop a more consistent position on this fundamental issue than he could have found in Kant. But at least at first his transcendental argument that the conception of ourselves as freely self-active agents governed by the self-legislated moral law is a condition of the possibility of consciousness itself seems, like Kant's thought on the issue, to be inexorably torn between the ideas that as free beings we are in fact necessarily governed by the moral law and that as both natural and rational beings we are necessarily governed by the ideal of the moral law but do not necessarily live up to it. However, Fichte might ultimately provide a way out of this dilemma, already hinted at in his account of moral feeling, by arguing that recognizing the moral law as an 'infinite' ideal that we necessarily strive to fulfil but to which we equally necessarily only asymptotically approach might be sufficient to satisfy the demands of the transcendental theory of self-consciousness without excluding the possibility of immoral actions and even of self-consciousness of them.

Fichte's attempt to deal with the problem of freedom goes on for many pages, and I will have to simplify it here. What is striking is that throughout the twists and turns of his argument, Fichte seems repeatedly torn between the same two positions between which Kant could never quite decide. On the one hand, Fichte often suggests that the free agent who can fully live up to the demands of morality and the agent who is played upon by natural forces that might appear to be the source for immoral choices are only two ways in which the same self appears, so that there can be no real conflict between

⁸ I have recounted the story of Kant's succession of attempts to deal with the problem of freedom in my *Kant*, 2nd edn (London: Routledge, 2014), ch. 6, and in 'Problems with Freedom: Kant's Argument in Groundwork III and its Subsequent Emendations', in Jens Timmermann (ed.), *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide* (Cambridge: Cambridge University Press, 2009).

them. Thus he writes that 'it is always I myself that determines me, and in no way I am I determined by the drive', and continues:

The ground for relating these predicates to one another is the following: although a part of what pertains to me is supposed to be possible only through freedom, and another part of the same is supposed to be independent of freedom, just as freedom is supposed to be independent of it, the substance to which both of these belong is simply one and the same, and is posited as one and the same. The I that feels and the I that thinks, the I that is driven and the I that makes a decision by means of its own free will: these are all the same. (SE 104)

This makes it sound as if there hardly can be any conflict between our natural drives and our free will, so that there can be nothing that can stand in the way of complete compliance with morality. As does this passage some pages later:

Are my drive as a natural being and my tendency as a pure spirit two different drives? No, from the transcendental point of view the two are one and the same original drive, which constitutes my being, simply viewed from two different sides. That is to say, I am a subject-object, and my true being consists in the identity and indivisibility of the two. If I view myself as an *object* completely determined by the laws of sensible intuition and discursive thinking, then what is in fact my one and only drive becomes for me my natural drive, because on this view I myself am nature. If I view myself as a *subject*, then this same single drive becomes for me a pure, spiritual drive, or it becomes the law of self-sufficiency. All the phenomena of the I rest solely upon the reciprocal interaction of these two drives, which is, properly speaking, only the reciprocal interaction of *one and the same drive with itself*. (SE 124–5)

On this account, human action cannot be understood as the product of two genuinely independent forces, natural drives (presumably to satisfy contingent inclinations) on the one hand, and a pure, spiritual drive to be moral on the other hand, either of which drive might but neither of which must dominate the other; the natural and the spiritual or moral drives of the human being are rather one and the same thing seen from two different points of view, so they cannot conflict with one another, and there would be, it seems, no circumstances in which the former could lead to an outcome contrary to what is demanded by the latter.

On the other hand, Fichte also insists that there must be room for moral failure, which can apparently be explained only by the assumption that our natural drives do not always coincide with our moral drive and sometimes get the upper hand over the latter. Thus he says that 'what drives our nature and determines our physical power need not be only the moral law itself. After all, we are also able to carry out immoral decisions' (SE 75). Thus, 'The law reason gives to itself . . . that is, the moral law, is not a law that it obeys necessarily, since it is directed at freedom' (SE 60): here Fichte not only maintains that the human agent does not necessarily act in accordance with the moral law, but even suggests that since the moral law is a law that commands freedom or self-activity, it would make no sense to think of it as a law that we are somehow necessitated to follow—the goal that this law commands us to realize would be undermined if we were somehow automatically determined to act in accordance with this law. But if

it is possible for me to find cause in my nature to carry out immoral decisions, then all of my natural drives and my pure spiritual drive to be moral cannot simply be the same thing under two descriptions or seen from two points of view: two descriptions of one and the same thing need not mention all the same facts about it, perhaps they need not mention any of the same facts, but they cannot be outright incompatible.

Sometimes Fichte seems to try to avoid this dilemma by the strategy that Kant had already tried in the third section of the *Groundwork*, that of allowing that there are two potentially conflicting sides to the self, but that only the drive to be moral reflects the 'proper' or 'authentic' self. Thus Fichte writes that

the natural drive, understood as a drive that is determined in a certain precise way, is contingent to the I itself. Viewed from the transcendental standpoint, it is the result of our own limitation. To be sure, it is indeed necessary that we be limited in some way or another, for otherwise no consciousness would be possible. But it is contingent that we be limited in precisely this way. In contrast, the pure drive is essential to the I; it is grounded in I-hood as such. For this very reason it is present in all rational beings, and whatever follows from it is valid for all rational beings.—Moreover, the pure drive is a higher drive, one that elevates me above nature with respect to my pure being and demands that I, as an empirical, temporal being, elevate myself above nature. (SE 135)

Fichte argues that a merely natural drive 'has no control over me, nor is it supposed to have any such control: I am supposed to determine myself utterly independently of the impetus of nature', but the fact that he repeats the words 'supposed to' suggests that he does not think it is automatic that the moral law is in control, so that even if the moral law represents the 'higher' part of our nature it does not represent the whole of it. Here Fichte does not claim that contingent natural drives and the moral drive are two sides of the same coin, instead allowing that the former might not be controlled by the latter, although they ought to be, because only the latter expresses our 'higher' and 'essential' self.

But if free self-activity that is not determined by any merely natural causal processes is supposed to be a condition of the possibility of self-consciousness, then this solution to the dilemma of freedom seems to raise a question about how we could even be conscious of merely natural drives that are not under the control of our freely chosen commitment to the moral law as part of our self, even if only a lower or contingent rather than essential part. And Fichte needs a solution to this question, because he has reasserted his commitment to the transcendental thesis that freedom is the condition of self-consciousness just a couple of pages previously:

I, however, am an I solely insofar as I am conscious of myself as an I: that is, as free and self-sufficient. This consciousness of my freedom is a condition of I-hood . . . a rational being is not possible at all without any consciousness of this freedom, and thus also not possible without the conditions for such freedom, and since one of these conditions is a consciousness of morality, a rational being is also not possible without such consciousness . . . Here we are claiming only that no human being could be absolutely lacking in any moral feeling. (SE 132)

If being conscious of myself as free and self-sufficient is a condition of self-consciousness or 'I-hood' itself, it might seem as if I could not even be conscious

of a part or aspect of myself that is resistant to morality, thus that the possibility of a non-moral part of myself might be a theoretically possible explanation for my immoral choice but not an explanation for which I could have any evidence. Fichte could avoid this problem, however, by holding that the awareness of *any* freedom, of 'any moral feeling' as the empirical manifestation of freedom, is sufficient for one to conceive of oneself as a self, rather than that an awareness of the freedom of *each* of one's acts is a necessary condition for the awareness of each of those acts. This assumption would explain why he is not worried about the possibility that one would be literally unaware or unself-conscious of any of one's natural drives that have not successfully been brought under control of one's rational commitment to the moral law, and is instead concerned only to maintain that as long as one is self-conscious one must be aware of the *call* of the moral law and of the *possibility* of living up to it, thus aware of oneself as *striving* to live up to the moral law. In order to be self-conscious, one must be aware of the moral law and strive to fulfil it, but there is no need for this striving to be fully satisfied, and indeed its complete satisfaction could be indefinitely or 'infinitely' remote. This is what Fichte himself presents as the resolution of the dilemma created by his making awareness of freedom the condition of the possibility of self-consciousness:

The causality of the pure drive must not disappear; for only insofar as I posit such a drive do I posit myself as an I.

We have arrived at a contradiction, which is all the more remarkable since what is contradictory is in this case . . . a *condition for consciousness*.

How can this contradiction be resolved? . . .

This can be comprehended only as follows: the intention, the concept that is involved in acting, aims at complete liberation from nature. But it is not as a consequence of our freely designed concept of the action that the action is and remains suitable to the natural drive; instead, this is a consequence of our limitation. The sole determining ground of the matter of our actions is [the goal of] ridding ourselves of our dependence upon nature, regardless of the fact that the independence that is thereby demanded is achieved . . . Consequently, the final end of a rational being necessarily lies in infinity; it is certainly not an end that can ever be achieved, but it is one to which a rational being, in consequence of its spiritual nature, is supposed to draw ceaselessly nearer and nearer. (SE 142)

For Fichte, the condition of the possibility of self-consciousness turns out to be an awareness of the necessity of *striving* to fulfil the moral law, not of ever fully succeeding in fulfilling it, but this awareness of striving is then sufficient to spread self-awareness over all of one's self, not just that part of it which is compliant with morality.

Fichte has to give up the claim that all of our natural drives and our pure spiritual drives are exactly the same thing just seen from two different points of view if he is to be able to appeal to a conflict between them in order to explain the possibility of immoral choices. But he does not have to give up the claim that being aware of the moral law is the condition of self-awareness itself if he is willing to allow that, once achieved through awareness of the moral law, self-consciousness can spread itself even

to immoral elements of our selves. If it is striving to be moral that makes us aware of ourselves as selves, that is compatible with failure to be fully moral.

Fichte's claim that the final end of a rational being is 'infinite' in the sense that it can *never* be fully achieved might seem to be a non sequitur, an inference from the mistaken supposition that an awareness of striving and thus of some opposition that always needs to be overcome is a *necessary* condition of self-awareness rather than an inference from the more plausible premise that even an awareness of striving to fulfil the moral law is a *sufficient* condition of self-consciousness. But, this mistake apart, perhaps he does after all suggest a way between the two horns of Kant's dilemma about freedom, and perhaps he does that precisely by keeping his argument a transcendental rather than a metaphysical or transcendent argument. That is, Kant's problem about the possibility of evil arose because he made the claim that the moral law is the causal law of the noumenal self, but Fichte introduces the moral law only as the condition of the possibility of self-consciousness, and this seems compatible with the assumption that we are not moral in every one of our acts, even while we may be fully aware of those acts. So perhaps in this area too Fichte points the way to a successful use of transcendental method in moral philosophy.

However, there would still seem to be one serious objection to the transcendental argument from the fact of our self-consciousness to the fact of our freedom. It does not seem that I could use the fact that self-consciousness implies awareness of my freedom to exclude the possibility that I (or anyone else) lacks freedom in at least some and perhaps morally crucial circumstances, because the connection between self-consciousness and consciousness of freedom could be a merely analytical proposition, sufficient to establish freedom with respect to as much of myself as I am conscious of but not sufficient to establish that there is no unfreedom in myself of which I am not immediately conscious. The proposition might establish that the limits of my self-consciousness are the limits of my freedom, but not that there are no limits to my self-consciousness and thus no limits to my freedom. The strategy of using the non-controversial fact that we are self-conscious to prove the unlimited reality of our freedom might seem as doomed as any other attempt to prove that we have unlimited freedom; it might instead only prove the limits of our self-consciousness.

But this might not be a bad outcome for a transcendental argument from self-consciousness to freedom after all: a Spinozist or a Freudian, after all, holds that one's freedom is limited precisely by the limits of one's self-awareness, and that the connection between freedom and self-awareness does not guarantee the unlimited extent of one's freedom but rather calls for an effort to extend one's self-awareness. I do not think that Fichte intended to reach this conclusion; on the contrary, his transcendental strategy of arguing that the idea of our freedom is the condition of the possibility of self-consciousness itself suggests that he at least started with the intention of proving the unlimited reality of our freedom. But he does seem to have at least backed into the

position that our freedom must take the form of an endless striving, not anything that is ever fully realized, and perhaps his own connection between self-consciousness and freedom shows what must be expanded in order to make progress in striving to be free, namely, self-consciousness itself. This would not be a bad moral for a transcendental theory of freedom.⁹

⁹ This chapter was written during my tenure as an Alexander von Humboldt Stiftung Research Prize Winner. I would like to thank the Alexander von Humboldt Stiftung for its generous support, and the Institut für Philosophie of the Humboldt Universität-Berlin and my Gastgeber, Rolf-Peter Horstmann, for their hospitality during this period.

Finite and Absolute Idealism

The Transcendental and the Metaphysical Hegel

Robert Pippin

I

Typical metaphysical questions include the following. Is the universe composed of dimensionless, windowless monads, or is it one infinite, eternal substance, or are there many substances but in real relations with each other? Is there free will or are all events determined? Is the mind a separate, immaterial substance? Is there an immortal soul, a God? Do abstract objects, like sets and universals and possible worlds exist? And so forth. These are not typical Hegelian questions and he does not show much interest in them. As he frequently makes clear in compliments to Kant, Hegel does not want to be enlisted in the tradition of rational dogmatism, or in any project that holds that we can settle questions about 'the furniture of the universe' by appeal to the light of reason or to tests of conceivability and inconceivability.

It turns out, though, that it can be quite misleading (I have discovered) to label Hegel somewhat casually a 'non-metaphysical' thinker, as Klaus Hartmann long ago suggested. For example, in my view Hegel held that agency was an achieved social status, ascribed in differing ways with differing scope over time; that assumptions about individuality behind methodological individualism in modern political philosophy were incompletely thought out; that acting freely involved the establishment of a determinate self- and other-relation; that explanations of the activities of some organic beings in complex social conditions would be inadequate if restricted to the material, or neuro-biological properties of such organisms. There is a perfectly good sense in which all these sorts of claims could and should be called 'metaphysical'. Hegel is purporting to tell us *what agency is*, *what individuality is*, *what freedom is*, and so forth, and he thinks of his answers as *distinctly philosophical*, not empirical or natural scientific. The key difference between these latter sorts of questions and the former sort comes down to how one understands Hegel's claim that his 'metaphysics' is primarily a

'logic'. (These three issues in particular have to do with the logical differences between and the relations between a philosophy of nature and a philosophy of spirit. Such a logic is primarily an account of the status and content of categories and a theory about explanatory adequacy, but it can have substantive results like those just listed.) No one in the history of philosophy had claimed anything like this before Hegel, although the ancients and the scholastics certainly had many things to say about categories. But for them, metaphysics was something different from and more important than such a category theory. In the following I propose to review the core issues in the logic–metaphysics relationship as Hegel understands it, and then discuss in detail a slogan recently used to describe Hegel's thinking about thinking and therewith his metaphysics. This is John McDowell's phrase, the 'unboundedness of the conceptual'.

The immediate temptation in trying to understand why Hegel thinks his logic is also a metaphysics is to take one's bearings from Kant, who inaugurated what a later commentator called a 'metaphysics of experience', and who himself said many things that sound like Hegel's Schellingean 'identity theory' about the relation between pure concepts and objects in the world. (Such as: 'The conditions for the possibility of experience [pure concepts] are at the same time the conditions for the possibility of objects of experience [constitute what could be an object]', A158/B197.) But this can be misleading too, as I have also found. It is clear that Hegel is not a transcendental philosopher, restricting the 'reality' which concepts constitute to some limited domain, the sphere of distinctly human experience, and inapplicable to things in themselves. (The subtitle of my chapter, echoing one by Richard Rorty, might have been: 'What do you say when they call your Hegel "a mere transcendental philosopher"?') So we need to start at the most basic issue and work up from there.

II

Throughout his life Hegel characterized his own position by partly invoking and appropriating, and partly criticizing, what he took to be the Kantian understanding of the relation between understanding and sensibility, concept and intuition.¹ All the passages clearly indicate that what Hegel is out to criticize is not the distinction itself, but the way Kant understands the nature of concept-intuition unity in knowledge claims. A 'mechanical' sort of application or inductive reflection, as he calls it, is what is being rejected in favor of what he calls an 'organic' understanding. But all the passages in Kant already have a dialectical and somewhat unstable form, as if already foreshadowing Hegelian logic. *Both* distinctness *and* necessary intertwining (inseparability in any claim to knowledge) are emphasized. This Kantian position raises two huge problems. The first is the issue of the right way to state the implications of the twin claims for any

¹ *Hegel's Idealism: The Satisfaction of Self-Consciousness* (Cambridge: Cambridge University Press, 1989); *HI* hereafter.

analysis of empirical knowledge. This will take us directly into the issues of the various myths (of the given and of the mental) that what was originally the Kant–Hegel disagreement have raised in contemporary discussions. The other is the issue of idealism; whether the inseparability claim as Hegel understands it, in his major difference from Kant, idealizes or relativizes to *us* any philosophical claim about objects.

I have said that the relevant passages in Hegel make clear that he is very much in agreement with Kant about the necessary *cooperation* of such elements in knowledge (thus accepting that there are such elements). But one should be careful. Interpreters who are interested in this line of thought have been portrayed several times as having Hegel ‘collapse’ the distinction between concept and intuition, and Michael Friedman has charged that the post-Kantian idealists ‘rejected’ outright Kant’s *distinction* between concept and intuition, and that they embraced a wholly self-determining *Vernunft* operating without empirical constraint.² (Friedman characterizes as ‘traditional’ what I would consider quite a bizarre thesis to attribute to anyone, that the ‘idealist doctrine that the world to which our thought relates is a creature of our own conceptualization.’ This ascription to Hegel of such a doctrine of intellectual intuition is one familiar way in which commentators understand Hegel’s ‘metaphysics.’³) But the passages are quite clear: Hegel never denied this *distinctness* claim, indeed he insisted on it (for example, in the passage before about an ‘organic’ and not ‘mechanical’ unity of such *different* epistemic dimensions of experience).⁴ What is true is that Hegel wished to stress *more*, make more out of, the organic unity or organic inseparability of such elements than Kant, where organic just means what it always has: that a severed hand is not any longer a *hand*, an intuited content considered separated from or in isolation from its role ‘inside’ the act of judging cannot be a contributory element to knowledge. This amounts to claiming that the ‘blindness’ of intuitions considered apart from conceptualization has different implications than Kant allowed, and changes what one can claim about a non-derived concept having an intuited content, being objectively valid.

This organic unity claim is the first manifestation of the claim that the conceptual is unbounded, that conceptual content cannot be understood as supplied ‘from without’ by epistemologically distinct intuitions, or is the Hegelian version of the familiar attack by Wilfrid Sellars on ‘the myth of the given.’ But it calls immediately to mind an understandable hesitation about the direction already suggested. The first is the worry that the unboundedness claim amounts to a kind of ‘intellectualism’, that understanding consciousness itself as an *activity* and purport as a *result* must mean that such activity must *be* apperceptive judging. Since there are obviously many sorts of relatively unreflective engagements that clearly do not fit such a model, the suspicion is that such a position greatly exaggerates the ‘intellectual’ dimension of experience.

² Michael Friedman, ‘Exorcising the Philosophical Tradition’, in Nicholas Smith (ed.), *Reading McDowell: On Mind and World* (Routledge: London, 2002), 25–57, esp. 33.

³ Friedman, ‘Exorcising the Philosophical Tradition’, 464.

⁴ *HI* 85.

But this notion of an unbounded conceptual articulation need not be committed to such an exaggeration. Consider the case of the practical domain. What gets attended to in praxis as salient, of ethical significance, what goes unnoticed in a well-functioning egalitarian society (eye color, race, gender, etc.), what occurrence raises a question, demands attention, what does not, who is taken to be of relevance to the moral community (humans, animals, the severely disabled), who is not (plants), and so forth can all be imagined to be of great and 'unreflective' weight in our practical world, some so deeply unreflective that it is hard even to imagine ever 'questioning' them, and yet it is highly implausible that such historically and culturally quite variant elements could be said to have any immediate, direct presence in our experience, as if pressing on our attention in themselves, 'on their own', from the 'outside'. A highly complex conceptual or normative interpretive framework is at work, without it being the case that such a being at work is a matter of explicit 'reflective endorsement', or the result of articulated moral evaluation.

III

This is all not, I take it, a revelation to anyone interested in Hegel; the opening argument of the *Phenomenology* famously starts us down that different, Hegelian path that will eventually lead to a much different understanding of the 'subject-object' relation than that typical in so-called 'reflective' philosophies like Kant's. But going further into the contrasting Hegelian claims about idealism immediately encounters two forbidding sorts of formulations. The first are frequent claims about not a concept-intuition relation but about *thought's self-negation*, and the second involve just as frequent formulations about the Concept *giving itself its own content*, the Concept being Absolute or in that term that McDowell has made well known, 'unbounded'. To be fair to critics like Friedman, this can certainly sound like we are talking about concepts unbounded *by* intuitions, that Hegel has rejected the 'discursivity thesis'⁵ according to which human thought can give itself no content but only categorize content provided 'from without'. What I would like to do in the following is to present a brief gloss on the first of the two issues (which, once placed in the context of the language developed by Kant's successors, is not as mysterious as it sounds), and then spend the rest of the time trying to understand what a concept giving itself its own content means, which is somewhat mysterious if it does *not* mean collapsing the distinction between concept and intuition, and so proposing a metaphysics in which the basic structure of reality is conceptual, a kind of neo-Platonic monist realism. Such a position, in the terms introduced here would mean insisting on not just inseparability in cognition between concept and intuition but actual indistinguishability.

⁵ See the account of the discursivity thesis in Henry E. Allison, *Kant's Transcendental Idealism: An Interpretation and Defense*, revised and enlarged edition (New Haven: Yale University Press, 2004).

Here is a typical statement about negation from his *Berlin Phenomenology*, as challenging to an interpreter now as it must have sounded then to his first readers.

The I is now this subjectivity, this infinite relation to itself, but therein, namely in this subjectivity, lies its negative relation to itself, diremption, differentiation, judgment. The I judges, and this constitutes it as consciousness; it repels itself from itself; this is a logical determination.⁶

There are scores of other passages throughout various works where Hegel appeals to this notion of the subject of thought, the 'I' that thinks, as a negative self-relation, a self-diremption, or original self-separating or self-repelling. And in a general sense this abstract formulation is not foreign; it is well known from Spinoza's insistence that all determination is negation (*omnis determinatio negatio est*). Determinate relation to an object or content is an exclusion or *negation of all that such a content is not*, a restricting or fixing of content that excludes. The peculiarity of the Kantian and post-Kantian formulation is that Hegel is not just saying that intending a content excludes as well as determines (he says rather that judging is a *self-negation*) and that such a fixing and negation is a *result*; that consciousness just is that *self-negating*. Since it is such a self-negating, any determinate take on the world is also implicitly and potentially self-transcending. Since intentional consciousness is a resolving and fixing of attention, such an activity is not wholly or uniformly 'positive'; it is also potentially self-negating in a broader sense; defeasible and reformulatable, and this must make some difference to the character of the determinate attending itself.

With respect to our problem, what I think Hegel is struggling to make clear here is that for him denying a separability to intuition does not damn us to reside within our own conceivings, as if shut up inside a 'world of our own conceiving'. When we are so attentive to this rather Fichtean point—that any conscious take on the way things are should be understood as potentially self-negating, not in any *matter of fact way* restricted, restrained or negated 'from the outside'—we see that no cognitive form of mindedness could ever be a matter of *being* merely 'positively' in a doxastic state, or merely *being* in any mode at all, as if a judgment could be a thing caused. To affirm is simultaneously to hold open the possibility that what one is affirming is not true, and holding this open in this way means that judging is always potentially self-transcending, aiming at the world as it is, not somehow confined as if as a matter of some fact within a worldview.

Arguments for relativism and sometimes for transcendental idealism often make this mistake, the mistake of thinking of thought or horizons of sense or modes of sense-making or conceptual schemes in this third-person way, as if something one can get trapped inside of unless something exogenous can 'break' through it. As it has been put in many contemporary contexts, one source of the confusion is the temptation to think in terms of conceptual schemes and a separable, otherwise neutral, non-conceptual content that is conceptualized by such a scheme. The temptation

⁶ G. W. F. Hegel, *The Berlin Phenomenology*, tr. Michael Petry (Dordrecht: Riedel, 1981), 2.

is to think of an in principle neutral or indeterminate content or world in itself the accessibility of which is a matter of applying a scheme to such a content and so ending up with something 'less' than the world in itself, but rather the world only as so finitely appropriated. Hegel is among the parties denying such a scheme-content distinction, although he is certainly not denying that there can be different, sometimes quite different, aspectual takes on the world. The point of this self-negating language is to distinguish this possible partiality of a 'shape of spirit' from the idea of some putatively radical, alternative conceptual scheme, and this view about the inherently possible self-negating aspect of such a 'shape' is meant to stress what Gadamer calls the 'openness' of linguistic horizons to each other.⁷ (Hegel is making a great deal more out of the fact that for him such an openness is not merely a feature of horizons; it is 'held open' actively.) This all allows a Hegelian distinction between a partial view of the world in some of its aspects, but intelligibly integratable with other partial aspects,⁸ and a contrary view of such aspectuality as due to the application of a scheme to a forever in-itself inaccessible content.⁹ Scheme and content might be inseparable in any take on the world, but that does not mean they are indistinguishable, as if such partiality could not be noted and corrected, expanded, as if content is wholly 'constituted' by the application of a 'finite' scheme. I want to suggest that this is the point of difference between a finite and an absolute idealism.

The point I am trying to make here is simply that this highly unusual language about negation and self-repelling is meant to reject this 'alternate scheme' picture, and to insist that the distinction between what we take to be the case and what is the case is *one we make*, in response to what we learn about the world, not an intrusion from outside that happens to us, whatever that could mean. (In his practical philosophy Hegel invokes the same sort of point to deny a Humean or any naturalist explanation of action. Insofar as something should be counted as an action of mine, a thing done intentionally, for the sake of a desire or need, it cannot be thought of as a body moved or just 'kicked' into motion by a somatic cause. Whatever I am undertaking must be intentionally *sustained* as well, and so to sustain it means always to be able to fail to

⁷ Cf. McDowell's discussion, especially in relation to the Friedman charge also discussed here, in 'Gadamer and Davidson on Understanding and Relativism', in Jeff Malpas, Ulrich Arnschwald, and Jens Kertscher (eds), *Gadamer's Century: Essays in Honor of Hans-Georg Gadamer* (Cambridge, MA: MIT Press, 2002), 173–93. Besides being right (in my view anyway) about the set of Friedman, Gadamer, and Davidson issues, McDowell also broaches the question of what we need to say is 'shareable' by a linguistic community in order for this mutual intelligibility and integration to succeed and suggests the beginnings of what I would regard as a Hegelian case for the indispensability of an 'I-We' relation beyond the 'I-Thou' priority argued for by Brandom and, in effect, by Davidson on the priority of idiolects.

⁸ On the issue of integration, see the discussion in chs 1 and 2 of A. W. Moore's *Points of View* (Oxford: Oxford University Press, 1997), esp. the statement of his Fundamental Principle on p. 21.

⁹ This obviously commits Hegel to a very difficult task. Not every aspectual view, or conceptual array organized around points of salience, mattering, concerns, and so forth are partial in *this* respect, integratable into a more comprehensive position, and Hegel needs to help us sort out what gets to be designated a 'Gestalt des Geistes' and what falls short. And contrary to Moore, *Points of View*, these partial points of view are not for Hegel integratable by 'simple addition'.

sustain it, and that failure is also not something that merely happens to me, as if a causal force is just extinguished. I either cancel or sustain a commitment to an end as I enact a deed, and once again (in the Introduction to the *Philosophy of Right*) Hegel describes this in the abstract language of self-negation or self-diremption.) Hegel also sometimes uses the misleading language of 'identification' to say this: that we both identify with, or endorse, an assertoric judgment or course of action, even while we have somehow held open its possible negation and so have not, in another sense, identified with it. For Hegel, this is on the way to saying that any determination (say empirical determination) of thought is a determination by thought, a self-determination or even potential self-negation, and this is why, for him, the inseparability claim is so important to stress. But we need to back up a few steps to untangle this progression of claims.

IV

Now, it is possible to cite a list of passages where Kant says, in effect, that 'Objects can appear to us without necessarily having to be related to functions of the understanding' (A89/B122; see also A90/B122; and B145). But the question of what Kant means when he claims it is possible to 'intuit an object' independently of concepts is not thereby settled. On the face of it Kant only seems to be repeating that the intuitional *aspects* of any object perceived cannot be attributed to the results of the understanding's determination; he is not saying that a cognitively significant pre-conceptual experience of an object is possible. (He often speaks of a synthesis of apprehension in intuition, and of reproduction in imagination, and at A120 insists that 'imagination is a necessary ingredient of perception itself'.) It is thus somewhat misleading to raise the issue in the contemporary terms of non-conceptual content, as some recent commentators have done. Kant is not really talking about non-conceptual or any sort of intuitional *content* in the passages at issue, but rather only about the non-conceptual, *formal* aspects of any *relation* to an object. Precisely because of this restriction, there are no indications he takes such items to be cognitively significant when considered in isolation. And no conceptual holist need affirm that reference must be fixed wholly conceptually, where conceptually is understood roughly as descriptively. There is a demonstrative use of concepts too.

But if we want to retrace the Hegelian path from these reflections, we need another component not prominent in McDowell. Indeed, given interpretations like Friedman's, this aspect of Hegel's position is by far the most important to notice when considering the question of what the denial of a strict separability between concept and intuition actually means or amounts to. For even though Hegel has in effect given up the Kantian strategy for demonstrating the objective validity of the categories, he still maintains that the very possibility of objective purport requires a conceptual projection of possible experience, the normative authority of which cannot be tied to an empirical

derivation (or empirical 'deduction' as Kant would say), but just thereby such authority still remains a question. (He also thinks that principles or norms for action are not in some way rationalized strategies for the satisfaction of desires and interests, nor are they formal legislations by pure practical reason. Yet the question of their normative status also remains.) So the issue of the authority or legitimacy of non-derived (and non-instrumental) norms, once this mind-world model changes from Kant's (or the genuine 'spirit' of Kant is emphasized) extends very far in Hegel. The main point now is that this shift leaves in place reflective questions about the status of the normative authority of concepts and principles understood in this way. Does that mean we are left with some (for Hegel quasi-psychological) claim about subjective indispensability, an enterprise of frictionless spinning? Or a metaphysical claim about the 'conceptual structure' of reality in itself? It should not, since the outcome of Hegel's take on the deduction is supposed to involve an altered way of seeing the 'subject and object' relation, such that interpretations like these will seem to have made several distorted assumptions. The course of his attempt to convince us of this and to illuminate this altered sense of the mind-world relation is the task of the *Phenomenology of Spirit*.

As the *Phenomenology's* Introduction also makes very clear, Hegel realizes that his own version of, let us say, the (*avant la lettre*) Sellarsian side of all this, creates its own distinct problem, the problem of grounding in some other way (other than by appeal to pure intuition, the separable form of all givenness, or empirically or pragmatically) the normative authority (what Kant called the 'objective validity', Hegel the 'reality' or *Wirklichkeit*) of non-derived, normatively constraining (or, in praxis, action-directing) elements in experience.¹⁰ In a general sense the 'experience' of consciousness contributes toward breakdown or loss of normative authority in what had functioned as empirically unchallengeable and something like Kant's problem about synthetic a priori judgments must be addressed, but now without Kant's account of the pure (separable) forms of intuition.

V

I would like now to draw some tentative and sketchy implications from this way of understanding the difference between finite and absolute idealism, where 'finite' idealism claims a dependence of thought not only on the existence of independent objects, but on some exogenous constraint and guidance, and where absolute idealism claims that all determination of thought is determination by thought or self-determining as a self-negating. (There is an image used by Kant, and nicely borrowed and adapted

¹⁰ If concept and intuition are radically inseparable, then not only is a pure or foundational empiricism excluded, but Kant's doctrine of transcendental formality is also being rejected and we will need another way to account for concept determinacy and (especially in Hegel) basic conceptual change. For the relevance of this issue to similar questions in Brandom's appropriation of Hegel, see my 'Brandom's Hegel', *European Journal of Philosophy*, 13 (2005), 381–408.

by Stephen Engstrom, for explaining the proper sense of dependence: that we should think of thought's dependence on receptivity the way we should think of a bird's dependence on the air for flight. The air and its resistance *enables* such flying; and we thus should not think of this dependence as the air or wind *buffeting* the bird about one way or the other."¹¹)

First a bit of review and reformulation. Someone who is taken to deny that there are unique representations with non-conceptual content that play a cognitively significant role in experience (that is, not a merely causal role) is often taken to deny that such representations could be said to 'provide' the content for thought. (How could they if there aren't such separable elements?) The natural inference is that such a denial must entail that thought must be said to 'fill out' such content on its own, to 'give itself content', what McDowell calls the 'unboundedness of the conceptual', what Hegel calls 'the self-determination of the Concept'. One hears many spirited attacks on such a claim, most of them, I would suggest, based on a very uncharitable reading of the claim itself and all of them reminiscent of what is just as often claimed about the Hegelian insistence on the inseparability of concept and intuition. One hears, for example, that since animals consciously perceive, but without the ability of wielding concepts, standing behind judgments, etc., and since we resemble animals in our sensory embodiedness, we must also rely in similar ways on cognitively significant but not apperceived non-conceptual content. Or that the mere fact that the richness and detail of our sensory experience can exceed the power of our current conceptual arsenal to discriminate means that non-conceptual content must play a cognitively significant role in experience. Or that veridical illusion and the persistence of Müller-Lyer appearances demonstrates a recalcitrantly non-conceptual, cognitively significant aspect of experience (or real 'modularity' in our mental lives). Or (incredibly) that someone committed to McDowell-Hegel holism would have to be committed to the claim that no increase in empirical knowledge, no addition to our conceptual arsenal, is *ever* possible because, putatively, the claim is that our current concepts wholly determine what is accessible in experience. Now of course, sometimes, when Hegel is discussing the Concept's self-determination, he can seem to put matters in a way that opens him to such epistemologically dualist, empiricist objections. It is no surprise to see hands thrown up in frustration at this point. But the frustration is premature.

As we have seen, both Hegel's claim about conceptual self-determination, and McDowell's about unboundedness, have suggested to several readers that McDowell has slipped back, despite his avowals, into a 'frictionless spinning' model. The same is said of Hegel (again by Friedman, and others). In both cases such charges are quite hasty, as if again inseparability is being confused with an indistinguishability claim. If one is careful about that distinction, then such Hegelian claims only first mean to

¹¹ Stephen Engstrom, 'Understanding and Sensibility', *Inquiry*, 49 (2006), 17. The Kant passage is from the first *Critique*, B8–9. ('The light dove, cleaving the air in her free flight, and feeling its resistance, might imagine that its flight would be still easier in empty space.')

insist that thought's relation to objects is not, as it were, 'secured' by receptivity, by the deliverances of sensibility, *alone*. The unboundedness claim is first cousin to the Davidsonian thesis that the only thing that can secure or support a belief is another belief, or that it is only because our uptake of the sensory world is already conceptually articulated that these deliverances can assume a justificatory role. (Indeed since the sensory world is an immediate sensory take *on the world*, the claim could just as easily be taken to be realist: reality itself is conceptual, offers up 'what could be taken to be the case'. And when Hegel speaks of reason finding itself in the world, of there being everywhere in the world, 'the Logos', he can be taken merely to be reiterating the founding principle of Greek rationalism: that to be is to be intelligible; there is nothing in principle unintelligible. Or in the *Tractarian* language McDowell sometimes invokes, thought does not 'stop short' of the world; a way of thinking about an object (a *Sinn*) is not an intermediary entity between us and the referents of thought; it is a way of seeing the world.) There is still plenty of substantive content and empirical guidedness in experience on such a picture. The claim is only, again, that thought's relation to such objects cannot be secured or even intuitionally pinned down, by the deliverances of sensibility alone. The broadest way to restate the point is simply that the domain of the normative—in this case what ought to be claimed—is autonomous. (This is Hegel's chief 'metaphysical point', even though it is not a point in a *substantialist* metaphysics.) Principles constraining what we ought to believe, what could count as a possible object of experience or what one ought to do are wholly independent of claims about how the mind works or what people generally do or what the received world determines us to think. Fichte appreciated this point in the deepest way and built his whole philosophy around it.

With all that said though, if the conceptual is unbounded, the normative domain autonomous, what is the non-metaphorical meaning of the notion of the Concept's 'self-determination' or 'unboundedness'? But that question depends on another. So formulated, is this sort of version of Hegel even an idealism? And if so, in what sense, if not a realist or metaphysical-in-the-rationalist-sense idealism?

VI

The most expansive summary of such claims is that *the forms of judgment, the forms of thought, are the forms of things, of objects and events*. (Not that they *correspond* one to the other; that would be realism. One is the other, as in identity philosophy and Kant's Highest Principle of Synthetic Judgments.) At this extreme altitude one is reminded of similar controversial claims by Wittgenstein in the *Tractatus*, as at 5.6, 'the limits of my language mean the limits of the world', something Wittgenstein provocatively calls 'the truth in solipsism', and expands in 5.61 as 'We cannot think what we cannot think; so we cannot think what we cannot say either.' Although Wittgenstein appears to be talking only about what Kant would call 'general logic', the point at issue is a broad one and has been put to all sorts of uses. There are parallels in the distinct uses by Wittgenstein

of 'my' and 'we' and Kant's 'subject of experience' and 'subjective' (and the 'I' in the 'I think' that must be able to accompany all 'my representations'). For both, these terms refer to nothing *in* the world, but express the limits of the world, set the limits of what could be a world. Any encounter with anything in the world would presuppose, could not 'discover', such a subject. The subject is not one more object in the world somehow containing all else.

For both as well, since this last point means that the basic statement of idealism involves no reference to an empirical psychological or actual social subject (is no species, as Bernard Williams notes, of any sort of 'Whorfian' claim about language and worldviews), or idealism is not invoked here as an *explanation*, the form of thought or the form of language does not *explain* 'why we experience the world as we do', the claim threatens to seem either a tautology or, at a deeper level, much more unusual; something, but not even a possible *claim* at all. The first danger is clear enough. The basic statement appears to say: that which we can understand and state, we can understand and state; that which we cannot, we cannot. (This danger is evident in interpretations of Kant as 'restricting' knowledge to our 'epistemic conditions'. Any view like this which is *not* a tautology threatens to introduce a substantive or empirical subject and thus a substantive or material or psychological notion of 'limit'. That Kant can give this impression is what Hegel most of all is objecting to in his famous attack on the notion of 'limit' or finitude.)

The latter possibility, that the basic statement is not a claim at all, but still *shows* us something, appears to be the way Wittgenstein understands it. Putting it this way reflects a response to a deep problem in any statement of a non-metaphysical idealism. One statement of such a post-Kantian idealism asserts a dependence of sorts (what sort being the heart of the matter) between the form of that to which our representations answer and some aspect of our representing capacities, or, in a version that raises the tautology problem, a dependence between the *form of objects-known* and the *form of knowledge*. Hegel is not fond of such dependence language and prefers his own 'identity philosophy' statements and so prefers an idealism claim according to which the conceptual is unbounded and self-determining. But in either case, it is obvious that this dependence (or identity) cannot itself be one of the objects to which representations answer. If it were so formulated it would be false. But we need to be able to explain our purchase on something like the worldliness of the world, the possibility of a world of experience, in a way that does not mistake such a target for something our representations could answer *to*, could be a feature of the world. Just calling this dimension transcendental does not help much. Hence the understandable emphasis on 'showing'.

But thus opens a potential disanalogy between Kant and Wittgenstein. How wide a disanalogy and what the relevance is for Hegel are challenging questions. For Wittgenstein, coming to understand what, say, 'comprehending the meaning of a term' amounts to *for us*, is not an empirical report on how we go on, not an element of a socio-linguistics. It is simply coming to understand what comprehending the meaning

of a term or a rule could be. (The Kantian parallel would be: all that being an object of our experience could be.) Even though Wittgenstein later seems to entertain the possibility of beings minded other than we are, his point seems to be to show ultimately that there couldn't (intelligibly) be beings minded other than us. If we insist: 'But the impossibility of entertaining such other-mindedness holds only by *our* lights, for us', then we have not understood what was just explained: that there is no we or I *in* the world 'for' which things are; that the point of introducing the notion of 'our' forms of thought is to help us see that there could be nothing else but 'ours', if *forms of thought*. The truth in solipsism, in a famous Wittgensteinian twist reminiscent of Hegel's style, is the truth of realism; the 'we' in Jonathan Lear's phrase, is a 'disappearing we'. Kant's idealism is a robust empirical realism; imagining an intuitive, not a discursive intelligence does not render our forms of thought 'limits' beyond which there is something in principle knowable, but not knowable by us. The world created by the divine intellect in thinking it is the same world as the world we know, even if God knows it in more dimensions and in a different way. (The main difference: he knows it all at once as it were; our knowledge is partial and infinitely additive.) So, in the commonsense way the issue is sometimes put: if the question is, 'How could we possibly assume that there can be no gap between "all-that-is-knowable" and "our capacity for knowledge"?', the answer is not a demonstration or a deduction that there could not be such a gap, nor is it to misunderstand the question as if it were about empirical capacities. (It is a ridiculous game, of no philosophical interest, to speculate about the possibility that the human brain may never to able to understand, say, the nature of consciousness.) The right response is to focus on the confusion implicit in the suggestion that there is some real referent of the 'our' in '*our* capacity for knowledge'. If it is a capacity for *knowledge*, it is not merely 'ours'. Or, in Rorty's apt phrase, the skeptical worry about what might be the world in itself, considered independently of any way we might know it, is a 'world well lost'.

But here the disanalogy (with Kant and ultimately with Hegel) begins. Wittgenstein clearly does not want the limits of language to be the sort of limit which has another side, a limit like a fence or a barrier. Yet a phrase like 'the limits of my language' does imply a *restriction* of some sort. That is why the Wittgenstein version of the basic claim is not a tautology, even if not a claim in the normal sense (not, perhaps it would be clearer to say, an *explanation* of the forms of things by appeal to the forms of thought; this is the same sense in which the claim that mindedness requires a 'spontaneity' is not pointing to a non-causal power in order to offer explanations of mental activities). There is a point at which nonsense *begins*, something we could not make sense of but can recognize as nonsensical. This restriction, however, suggests no unknowable world, is only available 'from the inside', as Williams puts it by 'finding our way around inside our own view, feeling our way out to the points at which we begin to lose our hold on it (or it, its hold on us) and things begin to be hopelessly strange to us'.¹² Strawson's term

¹² Bernard Williams, 'Wittgenstein and Idealism', in *Moral Luck: Philosophical Papers 1973–1980* (Cambridge: Cambridge University Press, 1981), 160. This is, I think, exactly the Hegelian point, and could be an epigraph for the *Phenomenology*.

has become justly famous: the 'bounds of sense' (a phrase Strawson most definitely did *not* think of as a version of idealism, a *limit claim*—that it isn't and needn't be is his whole point).

Kant does not seem to think of things this way and *does* seem to use the notion of a limit as a barrier with another side, for which he was famously taken to task by Hegel. (One has to straddle the limit, stand on both sides, to understand it as a limit in this sense. In which case it is not a limit in that sense.¹³) Given all of this sympathy by Hegel with these sorts of critiques of limit notions, does this mean that we should understand Hegel's 'idealism' to be as little a substantive claim as Wittgenstein's, a way of showing the disappearance of the relevance of any 'we'?¹⁴

It is when we face the issue of the determinateness of what are claimed to be candidates for the enabling 'forms of thought', and the unavailability of the Kantian separable forms of intuition, that a new form of 'instability', one might call it, emerges, the 'power of the negative' that forms the heart and soul of the *Phenomenology* (the 'pathway of doubt and despair'). That is, to take the quickest route to the issue in Hegel, if a condition for possible objective purport is some sort of projection of possibility, conditions which cannot be accounted for empirically or deduced by pure reason from the possibility of thought at all, normative constraints on what could be conceptual content at all, then we must also have some way of taking into account that the normative authority of such principles not only cannot be established once and for all by a deduction, but that this authority also can break down ('internally') and has broken down historically. And we must be able to do this without objectifying, psychologizing, or sociologizing such collective subjectivity. That is, the breakdown involves an experience of partiality and incompleteness, not anything like alternate conceptual schemes and so alternate worlds. Given the inseparability claim, the reassurance we can be said to require given this possibility and this fact cannot at all be the very general and vague reassurance that objects, considered independently of such conditions, can be said to fit or match what we require. But this does not mean that there is *no* problem to resolve, not at least according to many passages in the *Phenomenology's* Introduction. In what

¹³ See the lapidary formulation by Moore, *Points of View*, 119: 'At a more general level, we cannot represent limits to what we can represent. For if we cannot represent anything beyond those limits, then we cannot represent our not being able to represent anything beyond those limits.' I don't think Moore gets the relation between these issues and the Kantian idealist tradition in proper focus, because he (admittedly) passes over the important 'transcendent-transcendental' distinction in Kant. See *Points of View*, 122 n. 8. (By the *List der Vernunft*, though, the distinction gets a deep hold on his own enterprise in the last third of his book, when he distinguishes between a reflective level of analysis wherein his own Basic Assumption does not hold, and absolute representations are impossible, and a 'non-transcendental' level wherein the assumption does hold and absolute representations are possible. The contention with Kant and Hegel is over the claim that the former sort of knowledge is 'ineffable'. There is no greater opponent of ineffability in the history of philosophy than Hegel.)

¹⁴ There is still a lot to say here. If Wittgenstein is taken *sensu stricto*, he is talking about general logic in many of these passages, and pointing to the limits of what is logically expressible is not pointing to a real 'limit'. But more seems to be at stake especially in *Philosophical Investigations*. Cf. Williams's comments on how Wittgenstein's position is naturally driven to some sort of statement of transcendental idealism which must be false by its own lights ('Wittgenstein and Idealism', 163).

we might call 'normal' experience, within what Hegel names a 'shape of spirit', there are norms which cannot be questioned because they are the basis for the possibility of any questioning, norms which both Hegel and Wittgenstein say we are 'certain' of. That consciousness is direct and immediately presented with determinate objects which it can pick out and refer to indexically is not a theory or claim. It is more like a picture of what experience might be, what the mind-world relation is. As noted, this is not a claim about the mind-world relation, as if about another object in the world. That such a form of thought is the form of objects in such a context must function as a platitude. And Hegel 'examines' its sufficiency, he says, by 'watching', looking on, as an experience so shaped could be imagined trying to say what it knows. It cannot, and another picture is introduced.

One is tempted to say that this is a Hegelian response to a 'skepticism' problem and various schemes. Hegel was tempted to say it that way, and occasionally succumbed to such a temptation. But that is a misleading formulation of the issue, suggesting as it inevitably does a perspective 'outside' ordinary empirical and ethical claims, from which the very possibility of such claims can be established. But the *Phenomenology* remains phenomenological throughout, and this means that Hegel treats the way 'consciousness suffers violence at its own hands' as part of *what it is to have experience at all* (and that means to have a world at all), to be responsive to failures of practices of justification and legitimation, described as if 'from the inside', to stay with that image.

To be sure, in the Consciousness-Self-Consciousness-Reason chapters, Hegel is presenting an idealized picture of the education of consciousness about its own possibility, and so the self-negation is idealized, but the point towards which such education leads is a final corollary of sorts to the inseparability of mind and world that was the key point in Hegel's appropriation of Kant's deduction. This inseparability does not mean that transcendental logic, established by philosophical method, simply subjectively constitutes what the form of objects could be. The Hegelian direction, with respect to his infamous 'identity' claim, goes the other way. Inseparability for Hegel means that a logic interwoven in a form of life, a form of actual, historical life, cannot be rightly understood in abstraction from, separate from, the 'life' it regulates, and these forms or norms fail or break down in time, in some way lose their grip within such a form of life as a whole, such that all philosophy can be is 'its own time comprehended in thought'.

8

Is Hegel's *Phenomenology of Spirit* an Essay in Transcendental Argument?

Stephen Houlgate

In his famous article on the opening arguments of the *Phenomenology* Charles Taylor contends that the first three chapters of Hegel's book can be read as 'an essay in transcendental argument'.¹ The purpose of this argument, Taylor maintains, is to demonstrate 'the inescapable role of the concept' in human experience.² Similarly, Robert Pippin understands Hegel to be a transcendental philosopher who 'is, from the start, interested in the conditions of the possibility of knowledge'.³ Pippin's Hegel argues that knowledge and experience are possible only on the basis of certain conceptions or 'notions' of objectivity that are taken to be authoritative by (more or less) self-conscious subjects.⁴

There is much that I agree with in the interpretations of Hegel offered by Taylor and Pippin. My aim here, however, is to make one very simple point: *pace* Taylor and Pippin, Hegel's *Phenomenology* (1807) is not an essay in transcendental argument. Indeed, the distinctive character of phenomenology, as Hegel understands it, stems directly from his attempt to avoid the problems besetting transcendental thought.

In order to clarify the difference between Hegel's phenomenology and transcendental argumentation I propose first to examine an example of such argumentation in Kant. I am aware that neither Taylor nor Pippin claims that Hegel is a transcendental philosopher in precisely Kant's sense.⁵ Nonetheless, Hegelian phenomenology, as

¹ Taylor 1976, 151.

² Taylor 1976, 163.

³ Pippin 1989, 95.

⁴ See e.g. Pippin 1989, 114, 123.

⁵ See e.g. Pippin 1993, 57: '[Hegel's] rejection of both an empirical or naturalistic as well as a transcendental notion of subjectivity in favor of a notion of a subject of experience and action as necessarily *self-transforming in time* and necessarily *social*'.

Taylor and Pippin conceive it, retains certain problematic features of Kantian transcendental thought. By examining Kant's thought, therefore, we see most clearly what transcendental argumentation involves and what it is that Hegel, as I understand him, rejects.

1. Kant's Transcendental Deduction in A

In the transcendental deduction in the first (or A) edition of the *Critique of Pure Reason* Kant maintains that 'every intuition contains a manifold in itself'. He argues, however, that a manifold could not be represented *as* a manifold 'if the mind did not distinguish the time in the succession of impressions on one another'.⁶ The differences between our manifold representations do not, therefore, impose themselves directly upon us. We must be capable of intuiting differences *as* differences if they are to come explicitly to mind. What enables us to do so is our consciousness of time: we can explicitly differentiate one representation from another only because we can distinguish between what we see at one moment and what we see at another moment. Without this consciousness of differences in and of time, we would be incapable of discerning that we are given a manifold of *different* representations in intuition.⁷

The premise on which Kant bases his conclusion is this: if all we were aware of were one undifferentiated moment, whatever we intuit would itself remain for us an undifferentiated unity. As Kant puts it, '*as contained in one moment* no representation can ever be anything other than absolute unity' (CPR 228–9/A99). Our awareness of empirical difference (and indeed of spatial difference) thus depends completely on our ability to discern different moments *in time*.

Yet this is not the whole story: for we can discern differences between representations only if we can also hold those different representations *together*. If we were to perceive one thing now and another thing later, but these two perceptions were to remain completely separate and unrelated, we would not be able to detect that they are different *from one another*. Becoming conscious of a manifold of representations *as* a differentiated manifold thus requires not only that we explicitly distinguish the representations in time, but also that we hold them together in one intuition. In order to do this, Kant argues, 'it is necessary first to run through and then to take together this manifoldness, which action I call the *synthesis of apprehension*' (CPR 229/A99). Such synthesis must be empirical: the mind must be able to hold together different empirical representations, if it is to discern the difference between, for example, red and green. But synthesis must also be pure and a priori: the mind must be able to hold together moments of time *as such*, if it is to differentiate what is intuited now from what is intuited later. This pure, a priori synthesis of apprehension is the first

⁶ Kant 1997, 228 (A99). Further references to Kant's first *Critique* will be given in the main text in the following form: CPR 228/A 99.

⁷ See Longuenesse 1998, 37.

transcendental condition of differentiated intuition—and thus of experience—identified by Kant.⁸

This synthesis of apprehension is, however, inseparable from the synthesis of reproduction. This is because one holds together successive moments in time by uniting the present moment with moments that have passed; but one can do this only if one can *reproduce* past moments as one moves from one present moment to another. As Kant explains,

if I draw a line in thought, or think of the time from one noon to the next, . . . I must necessarily first grasp one of these manifold representations after another in my thoughts. But if I were always to lose the preceding representations . . . from my thoughts and not reproduce them when I proceed to the following ones, then no whole representation and none of the previously mentioned thoughts, not even the purest and most fundamental representations of space and time, could ever arise. (CPR 230/A102)

The condition of being able to hold together, or 'apprehend', different moments in time is thus the ability to reproduce such moments through the empirical and transcendental imagination as one moves through time.⁹

Kant immediately points out, however, that 'without consciousness that that which we think is the very same as what we thought a moment before, all reproduction in the series of representations would be in vain (*vergeblich*)' (CPR 230/A103). In such a case we would reproduce one moment at a later moment, but we would not recognize that what we are doing is *reproducing* in the present a moment from the past. Reproduction would thus be 'in vain' because it would not bring with it any consciousness that we are holding together the present moment and a different, past moment that has been *retained*.

For us to recognize that 'that which we think is the very same as what we thought a moment before', we need to grasp the identity that that reproduced item preserves through time; and we grasp this identity in a *concept*. Indeed, Kant claims that we need to understand the concept not only of the particular item that is being reproduced, but also of the whole that is thereby being apprehended. The concept, Kant tells us, is 'this one consciousness that unifies the manifold that has been successively intuited, and then also reproduced, into one representation' (CPR 231/A 103). Empirical concepts and transcendental concepts—or categories—are thus further conditions of experience.¹⁰

Concepts involve a consciousness of identity or unity in the synthesis of the manifold of intuitions. They also presuppose the thought of the identity or unity of the *consciousness* that undertakes this synthesis: for if that consciousness had no thought of its own unity—if it did not think of *itself* as a unified field of representations—it could not become aware of any unity among its representations. This thought of the unity of consciousness, Kant maintains, is not given with our spatio-temporal intuitions, but

⁸ See Longuenesse 1998, 38.

⁹ See Longuenesse 1998, 41.

¹⁰ See Longuenesse 1998, 45.

is a further transcendental condition of experience. Indeed, for Kant, it is the ultimate 'transcendental presupposition' of experience. This presupposition is, of course, '*transcendental apperception*' or transcendental self-consciousness (CPR 232/A107).

Such transcendental apperception does not take the form of a distinct consciousness of oneself as an 'I'. The explicit and distinct consciousness of oneself as an 'I' is what Kant calls (in the second (or B) edition of the *Critique of Pure Reason*) the '*analytical unity of apperception*' (CPR 247/B133). This is the consciousness of myself, expressed in the judgement 'I think', that must be able to (but does not have to) accompany all my representations (CPR 246/B131).¹¹ The transcendental apperception that is the ultimate presupposition of experience is subtly different from this: it is not the thought of myself as something *distinct* from my other representations, but is the 'original and necessary consciousness of the identity of oneself' in the synthesizing of those representations (CPR 233/A108). That is to say, it is the a priori consciousness of oneself as the identity or unity *in which* one's representations can be synthesized. Indeed, it is the a priori consciousness of oneself as the synthetic unity—or unified *field*—of one's possible representations. This consciousness or thought of oneself as the unity in which one's representations can be synthesized (or as the synthetic unity of one's possible representations)—which Kant in the B edition calls the '*synthetic unity of apperception*'—is what makes possible unified empirical experience and also the explicit consciousness of myself as an 'I' (or the '*analytical*' unity of apperception) (CPR 246–7/B131–5).

Now my aim in this section is not to examine Kant's transcendental deduction in detail or to assess its validity or coherence. My aim is simply to highlight certain aspects of Kant's transcendental procedure.

First of all, Kant starts out in the A deduction from two definite assumptions about intuition and representation. These are *philosophical* assumptions and are not identified by Kant as ones that non-philosophers would automatically share. The first is that 'every intuition contains a manifold in itself', and the second is that '*as contained in one moment* no representation can ever be anything other than absolute unity' (CPR 228/A99). The first claim is unlikely to meet with resistance from non-philosophers, but the second is far from uncontroversial. Furthermore, it is unclear how one would prove such a claim. Yet this unproven assertion plays a significant role in Kant's developing argument: for it underlies his claim that the mind can become explicitly aware of *differences* in the intuited manifold only by comparing the content of one moment in time with that of another, subsequent moment. This idea that the different aspects of the manifold must be brought to mind *one after another* then makes it necessary for Kant to argue that the mind must bring these elements together as different through the activity of reproductive synthesis.

¹¹ See Allison 2004, 164.

At the start of the B deduction, the claim that differences in what is intuited are brought to mind one after another plays no obvious role. Nonetheless, the B deduction rests on its own assumptions about sensation and intuition: namely that 'the combination (*conjunctio*) of a manifold in general can never come to us through the senses, and therefore cannot already be contained in the pure form of sensible intuition' (CPR 245/B129–30). This in turn supports the claim that combination or synthesis is an act of spontaneity undertaken by the mind. In this respect, therefore, despite starting from a somewhat more general premise, the B deduction overlaps closely with the A deduction.

The second thing to note about Kant's transcendental procedure is that it establishes conditions of the possibility of experience that are understood to be necessary *by the philosopher*. The ordinary, non-philosophical mind is not directly aware that it is carrying out transcendental synthesis when it perceives empirical objects, but simply finds itself confronted with such objects. The transcendental philosopher argues, however, that the mind *must* carry out such synthesis if the possibility of empirical experience (and of the objects of such experience) is to be explained. Transcendental reflection, for Kant, thus does not involve any immediate introspection; it entails, rather, working out rationally what the conditions of experience must be. This is especially important to bear in mind when considering Kant's concept of transcendental apperception.

The analytical unity of apperception is the direct consciousness of myself as an 'I' that is expressed in the judgement 'I think'. As we know, Kant states that this judgement must be able to accompany all our representations but does not actually have to accompany them (CPR 246/B131). That means that, for Kant, I can experience objects without always being reflectively aware *that* I am experiencing them. The transcendental condition of my being able to experience such objects, however, is transcendental apperception or the synthetic unity of apperception. This is a form of self-consciousness, or self-thinking, that does *not* involve the direct consciousness of myself as an 'I'. It is, rather, the a priori consciousness of oneself as the unity in which one's representations can be synthesized (or as the synthetic unity of one's possible representations).

Since such apperception is the *transcendental* condition of experience, it is not a self-consciousness of which I can ever be explicitly aware in my ordinary experience (in the way I can be conscious of myself as an 'I'). Rather, such transcendental apperception is *presupposed* by the philosopher as the necessary condition without which a unified experience of manifold objects could not be explained. To put this another way: I can never catch myself in the act of transcendently apperceiving, but the philosopher argues that at some level I must be conscious of myself as the synthetic unity of my possible representations, for otherwise—given the assumptions he makes about the manifold of intuition—he could not explain how it is possible for us to have the experiences we have (or at least that he takes us to have).

2. Hegel and (Quasi-)Transcendental Thought

Having now considered briefly what is characteristic of Kant's transcendental thought, I wish to argue that Hegel's *Phenomenology* is not an essay in Kantian transcendental argument in either a strict or a looser sense. Hegel's project is of a quite different kind. This is made clear in the Introduction to the *Phenomenology*.

At the beginning of his Introduction Hegel famously criticizes what he calls 'the natural assumption that in philosophy, before we start to deal with its proper subject-matter, viz. the actual cognition of what truly is, one must first of all come to an understanding about cognition, which is regarded either as the instrument to get hold of the Absolute, or as the medium through which one discovers it'.¹² Hegel's criticism is twofold. First, this so-called 'natural' assumption leads to skepticism concerning our ability to know the truth (because the 'instrument' or 'medium' will always be thought to get in the way). Second, it is, indeed, no more than an *assumption* that incorporates (or rests on) the further assumption that words such as 'absolute', 'cognition', 'objective' and 'subjective' all have a meaning that is generally familiar (*PS* 48/70–1).

After listing and bemoaning the unwarranted assumptions that have come to seem 'natural' to philosophers, such as Locke, Kant, and Reinhold, Hegel suggests that the proper thing to do might be just to set such assumptions to one side and get on without further ado with the job of knowing what there is. 'Instead of both-ering to refute all these ideas', Hegel writes, 'we could reject them out of hand as adventitious and arbitrary. . . . We could . . . simply spare ourselves the trouble of paying any attention whatever to such ideas and locutions' (*PS* 48/70–1). I have argued elsewhere that this is precisely what Hegel does in his *Science of Logic*: he sets all familiar assumptions about thought and being to one side and focuses on the utterly indeterminate thought of pure being. In this way, he hopes to be able to determine the true nature of being and of thought while taking as little as possible for granted about either.¹³ In the *Phenomenology*, however, Hegel takes a different course. He points out that speculative philosophy faces another opponent apart from Lockean and Kantian philosophers, namely ordinary or 'natural' consciousness.

It is not always recognized that the so-called 'natural' assumption attributed by Hegel to certain philosophers is not shared by ordinary, natural, non-philosophical consciousness. The philosophers Hegel has in mind at the start of the Introduction draw a *threefold* distinction between 'subject', 'object' and 'medium' (or 'instrument') of cognition. Ordinary consciousness, by contrast, draws a simple *twofold* distinction between itself and the object it knows (see *PS* 47, 52/70, 76). Ordinary consciousness thus takes itself to be in direct contact with the things it knows, and does not fear that

¹² Hegel 1977, 46; Hegel 1970b, 68. Further references to the *Phenomenology* will be given in the main text in the following form: *PS* 46/68 (with the page number from the English translation given first).

¹³ Houlgate 2006, 29–53.

it is being held at one remove from such things by any 'instrument' or 'medium' of cognition.¹⁴

Hegel regards the threefold distinction drawn by philosophers, such as Kant and Reinhold, as arbitrary and unwarranted. Philosophers, however, should be thoroughly self-critical beings and not just rest their theories on arbitrary principles and distinctions. Hence it is quite proper simply to set any such distinctions to one side and start philosophy with the indeterminate thought of pure being, as occurs in the *Science of Logic*.¹⁵ The twofold distinction drawn by ordinary consciousness is different. As we see later in the Introduction to the *Phenomenology*, it is a fundamental and quite evident feature of ordinary experience: all of us in our everyday comings and goings draw a distinction between ourselves and the things 'over there' of which we are conscious. Furthermore, it is not incumbent on ordinary, non-philosophical consciousness—as it is on philosophical understanding—to be thoroughly self-critical. This is not to say that such consciousness should be thoughtless, but it gets on with life by trusting in the distinctions it makes and the things it experiences and not calling everything into question in the manner of Descartes. Hegel recognizes, therefore, that the speculative philosopher cannot expect ordinary consciousness simply to set to one side its favoured assumptions about things in the way one expects philosophers to do so. He also recognizes that, as a consequence of their quite proper adherence to their own point of view, non-philosophers will see a certain arbitrariness and perversity in speculative philosophy (*PS* 15/30). This is part of what Hegel means by saying that 'Science, just because it comes on the scene, is itself an appearance' (*PS* 48/71): from the perspective of those outside of philosophy, speculative philosophy or 'Science' is an unjustified *phenomenon* with no special claim on our ordinary, non-philosophical attention.

At this point in the Introduction Hegel's focus switches away from the relation between speculative philosophy and Kantian-Reinholdian 'representational' philosophy. Indeed, nothing more is said about the latter philosophy or about the so-called 'natural' assumption with which Hegel opened his text. Hegel now turns his attention to the following question: what stance should speculative philosophy adopt towards ordinary, *non-philosophical* consciousness? It is in the course of considering this question that Hegel briefly sketches and rejects what one might call a 'quasi-transcendental' approach to ordinary consciousness. On the basis of Hegel's brief sketch I hope to

¹⁴ Pippin, for example, appears to miss this point when he writes that 'the assumption that generates a critical skepticism—a possible gap between our "Notions of objects" and objects in themselves (particularly our *a priori* Notions or categories)—is precisely the assumption that Hegel attributes to natural consciousness, the subject of the *PhG*' (1989, 111). It is true that Hegel thinks that natural consciousness distinguishes its objects from itself, but he does not claim that this distinction generates a critical skepticism. Such skepticism is generated by the assumptions that certain *philosophers* take to be 'natural', namely that cognition is an instrument or medium, that there is a difference between ourselves and this cognition, and that 'the Absolute stands on one side and cognition on the other, independent and separated from it' (*PS* 47/70). In contrast to this, natural consciousness—especially in its first form of sensuous certainty, but also in the form of perception—takes itself to be in direct contact with its objects.

¹⁵ See Hegel 1999, 69–70, 72.

indicate why Hegel also rejects Kant's own more properly transcendental approach (and why he would reject the interpretations of his *Phenomenology* provided by Taylor and Pippin).

Hegel writes that 'when confronted with a knowledge that is without truth, Science can neither merely reject it as an ordinary way of looking at things (*eine gemeine Ansicht der Dinge*), while assuring us that its Science is a quite different sort of cognition . . .; nor can it appeal to the ordinary view for the intimation (*Ahnung*) it gives us of something better to come' (PS 48–9/71). This passage suggests two stances that can be (but should not be) adopted by speculative philosophy towards ordinary consciousness.

The first is that of simple, direct rejection. Such rejection might be immediate or might follow extensive argument; the significant point is that this stance involves effectively saying to ordinary consciousness: 'you are in the wrong, and speculative philosophy is in the right'. This is a stance that many philosophers adopt towards non-philosophical (and philosophical) positions with which they do not agree. It is apparent every time a philosopher says to someone else: 'what you say is manifestly false'. In Hegel's judgement, however, this is not the appropriate way for philosophy to deal with ordinary consciousness. Whenever I reject another's position directly in this way, my rejection of that position ultimately rests on the assertion that *I am* right. This is the case, however complex and subtle I take my intervening arguments to be: underlying such arguments is always the assertion that they *are* subtle arguments even if *you* cannot see this. Yet such an assertion on my part invites the counter-assertion from the other that he or she *is* right. If speculative philosophy, or Science, were to confront and reject the standpoint of ordinary consciousness in this way, therefore, 'Science would be declaring its power to lie simply in its *being*'—in the claim that it *is* in the right—but ordinary consciousness 'likewise appeals to the fact that it *is*, and *assures* us that for it Science is of no account'. As Hegel points out, however, '*one* bare assurance is worth just as much as another' (PS 49/71). The direct, confrontational approach thus confers no advantage on speculative philosophy, and does not put the latter in a position to persuade ordinary consciousness of its merits. In fact, such a direct approach turns speculative philosophy and ordinary consciousness into mirror images of one another: each one asserts that it *is* right and that the other simply *is* wrong. Hegel thus dismisses this approach as inappropriate for speculative philosophy that wants to demonstrate to ordinary consciousness, rather than just assert, that its position is to be preferred.

The second stance that Hegel considers is subtly different from the first. In this case, speculative philosophy does not assert itself *against* ordinary consciousness, but 'appeal[s] to whatever intimation of something better it may detect *in* the cognition that is without truth, to the signs (*Hinweisung*) which point in the direction of Science' (PS 49/71, my emphasis). This, I believe, is the stance that Hegel himself adopted towards Kant in his 1802 essay, 'Faith and Knowledge' (*Glauben und Wissen*). In that essay, Hegel undertakes a close examination of several of Kant's central concepts, including those of the productive imagination and the synthetic unity of apperception. What

interests him about these two particular concepts, however, is not so much the distinctive role they play within Kant's own critical system, but rather the fact that they prefigure the true, speculative idea of reason as the 'identity of opposites'.¹⁶ Despite being profoundly critical of Kant for remaining within the narrow ambit of Lockean thought and restricting himself to examining finite understanding,¹⁷ Hegel thus detects within Kant's limited philosophy clear intimations of his *own* idea of true reason. Indeed, he goes so far as to claim (to the consternation of Kantians) that Kantian imagination 'is nothing other than reason itself', albeit imperfectly expressed.¹⁸

Note that this approach to others does not entail countering their errors with a truth that falls quite outside their understanding. It does not involve simply saying to the other 'you are wrong and I am right'. It finds the truth hidden, or imperfectly disclosed, within the perspective of the other itself. Such an approach I shall call 'quasi-transcendental'. This is because it shares an important feature with a more properly transcendental approach: namely, it establishes its superiority *indirectly* through reflection on what is entailed by the position of the other.

Properly transcendental reflection would examine ordinary consciousness and uncover the hidden conditions of such consciousness; and by uncovering in this way conditions that consciousness itself cannot see, such reflection would demonstrate its superior insight. The 'quasi-transcendental' reflection described by Hegel, by contrast, does not merely disclose the 'conditions' of consciousness. It proves its superiority by discerning in consciousness intimations of 'something better' than anything explicitly acknowledged by consciousness, intimations of the truth that will be fully articulated by speculative philosophy itself. Transcendental and quasi-transcendental reflection are thus subtly different. Nonetheless, they have this in common: neither simply confronts ordinary consciousness with the claims of philosophy and expects consciousness to give way, but both prove their superiority by seeing more in consciousness than consciousness itself can see.

Just as Hegel in the *Phenomenology* rejects the first—direct, confrontational—approach to ordinary consciousness, so also he rejects the second—'quasi-transcendental'—one. Neither is the appropriate stance for speculative philosophy to adopt. Furthermore, Hegel rejects the second approach for the same reason as he rejects the first: 'for one thing, [philosophy, or Science] would only be appealing again to what merely *is*; and for another, it would only be appealing to itself' (PS 49/71). Yet how can this be, if this second approach proceeds not by arguing directly for its own position, but by reflecting on what is entailed by the position of ordinary consciousness?

The problem, as Hegel sees it, is that in adopting this second, quasi-transcendental approach, philosophy finds in ordinary consciousness only what *philosophy itself*

¹⁶ Hegel 1970a, 305.

¹⁷ Hegel 1970a, 304.

¹⁸ Hegel 1970a, 308. See also Houlgate 2005b, 152–8.

can see there. In so doing, philosophy certainly finds truth in consciousness, but that truth is not apparent *from the perspective of consciousness itself* and in that sense does not belong to consciousness. It is a truth that philosophy alone, not ordinary consciousness, asserts to be present in consciousness. At first sight, quasi-transcendental reflection might look as though its attention is essentially directed away from itself to the perspective of the *other*. Yet in fact such reflection remains firmly enclosed within its own perspective and reaches—or confirms—insights that only it, and not the other, can comprehend. That is to say, quasi-transcendental reflection is always referred *back to itself* by the other on which it reflects, and in that sense can be said to ‘appeal’ only to itself in its understanding of the other.

We can now see why the Hegel of the *Phenomenology* would also reject a properly transcendental approach to ordinary consciousness. He would object to the fact that in adopting such an approach philosophy would identify conditions of ordinary consciousness which only *philosophy* can discern. In so doing philosophy would demonstrate its superiority over ordinary consciousness by simply appealing to its *own* philosophical insight. It would thus beg the question against such consciousness through the very method of its procedure. The transcendental approach to ordinary consciousness looks initially as though it would be preferable to the confrontational approach, because it does not set itself from the start *against* such consciousness. It does not try to show that it is right and the other is wrong. It starts, rather, *with* the position of ordinary consciousness itself, and seeks by reflecting on such consciousness to understand the conditions under which such consciousness can be what it is. It appears therefore to be a more generous, more understanding, and less aggressively superior approach to the other than the confrontational one. On closer inspection, however, the transcendental approach is not quite as far from the confrontational one as it initially appears: because the transcendental philosopher appeals to his own insights and own convictions as much as the confrontational philosopher does. The conditions of ordinary consciousness that are understood to be necessary by the transcendental philosopher are not recognized to be such *by ordinary consciousness itself* but are disclosed—behind the back of such consciousness—by the superior understanding of the philosopher. In this respect the transcendental philosopher appeals to *himself* and his *own* insight in his encounter with ordinary consciousness, as much as the confrontational philosopher does.

Yet is not Kantian transcendental argumentation more subtle than that? Does Kant not develop sophisticated arguments to show that, given the character of intuition, cognition must be subject to certain definite conditions? Yes indeed; but Kant’s arguments—undeniably subtle though they are—still disclose conditions of knowledge and experience of which non-philosophical consciousness remains unaware, and, as I have argued, they rest on specific *assumptions* that ordinary consciousness does not automatically share (such as that we can intuit different elements only one after another). From that point of view, Hegel’s complaint still holds true: in adopting a

Kantian transcendental approach to ordinary consciousness, philosophy cannot but appeal to itself in the form of its own insight and assumptions. It does not truly demonstrate its superiority over ordinary consciousness, therefore, but *asserts* it on the basis of the assumptions it makes (about the difference between understanding and sensibility, and so on).

Note, by the way, that something similar is true of Charles Taylor's Hegel. By Taylor's own admission, his Hegel 'start[s] from undeniable characteristics of experience'. These characteristics are evident to the philosopher, but not to ordinary consciousness itself. Indeed, 'they must be independent', that is, 'not derivable' from the model of experience adopted by consciousness, for otherwise they could not be used to 'contradict [that model] and show it to be impossible'.¹⁹

'Consciousness', Taylor writes, 'is an object that lives in relation to a model of itself': it is shaped 'by a certain *idea* of what experience is'.²⁰ Hegel's project in the *Phenomenology* is to demonstrate in a series of interrelated analyses the inadequacy of the model of itself and its experience that consciousness forms for itself. Hegel demonstrates this inadequacy by showing that, whenever consciousness tries to realize its model, 'effective experience violates it'.²¹ Taylor concludes, however, that 'if we are to show that the model is not just unrealized in a given case, but cannot be realized, we have to be able to identify some basic and pervasive facets of experience *independently* of our model'.²² It is only in relation to these facets of experience, which are presupposed by the philosopher, though not by the shape of consciousness under examination, that the model of experience adopted by consciousness can be shown to be inadequate. Taylor's Hegel thus rests his 'transcendental' argument on what Taylor (and presumably Taylor's Hegel) openly acknowledges to be assumptions that are independent of the perspective of consciousness itself.

The central assumption made by Taylor's Hegel about experience and knowledge is that 'to know is to be able to say'.²³ As Taylor notes, Hegel's procedure in the opening chapter of the *Phenomenology* is thus similar to Wittgenstein's: 'he challenges sensible certainty to *say* what it experiences'. Taylor continues: 'the underlying principle is the same, viz. that if this is really knowledge, then one must be able to say what it is, and this is (here as with Wittgenstein) the starting point of what we called above a transcendental argument'.²⁴ Sense-certainty, however, proves unable to say what it knows without going beyond the sheer immediacies of which it takes itself to be aware and subsuming them under concepts. In this way, Taylor's Hegel shows, 'the attempt to say will contradict the basic requirements of sensible certainty, will take us beyond its defining limits, and hence it will stand self-refuted'.²⁵ The word 'self-refuted', however, is really out of place here, since sense-certainty is not refuted purely by its own model of experience. It is 'refuted' by the failure of that model to survive the challenge,

¹⁹ Taylor 1976, 160.

²⁰ Taylor 1976, 158–9.

²¹ Taylor 1976, 159.

²² Taylor 1976, 160, my emphasis.

²³ Taylor 1976, 156.

²⁴ Taylor 1976, 162.

²⁵ Taylor 1976, 162.

addressed to consciousness by Taylor's Hegel, to say what it means. This challenge is made because Taylor's Hegel, though *not* sense-certainty itself, takes it for granted as 'the basic starting point that to know is to be able to say'.²⁶

The lesson that Taylor's Hegel draws from sense-certainty's failure is that all knowledge is mediated by concepts or universals.²⁷ Such concepts can thus be said to constitute the transcendental conditions of any genuine knowledge: they are that without which true knowledge is impossible. It appears from Taylor's account that *experience* teaches sense-certainty itself that we can hold the particular before our gaze 'only by subsuming it under a universal'. Furthermore, Taylor maintains that Hegel's own insight into the impossibility of unmediated knowledge of particulars and into the necessary role of concepts 'reflects our experience', and indeed is based on that experience.²⁸ As we have seen, however, the experience of sense-certainty leads to the conclusion that concepts or universals constitute the transcendental conditions of knowledge only because Hegel confronts sense-certainty with the *independent* principle that in experience one must be able to say what one knows. The insight reached by sense-certainty and by Hegel into the conditions of knowledge thus rests on a philosophical assumption that *Hegel* makes, but that is not shared by sense-certainty itself. To that extent Hegel's conclusion is very much his own philosophical conclusion, even though it 'reflects our experience'. It is not a conclusion that sense-certainty would have reached by itself had it been left to its own devices.

In this respect, Taylor's Hegel is close to Kant: he sets out what the *philosopher* recognizes to be the transcendental conditions of experience, and he does so by proceeding from certain assumptions about experience and knowledge that are made by the *philosopher* but not shared by ordinary, non-philosophical consciousness. In so doing Taylor's Hegel, like Kant, does precisely what Hegel in the Introduction to the *Phenomenology* says the philosopher should not do: he achieves a position of superior wisdom by appealing ultimately to *himself* (in the form of his own assumptions about cognition and his own insight based on those assumptions). As we shall see later, Pippin's Hegel does something very similar.²⁹

²⁶ Taylor 1976, 156.

²⁷ Taylor 1976, 163.

²⁸ Taylor 1976, 168.

²⁹ My aim in this chapter is to point to certain features of transcendental argumentation as *Kant* conceives it and to show that, *pace* Taylor and Pippin, Hegel's phenomenology is not an essay in transcendental argumentation in the broadly Kantian sense (or an exercise in quasi-transcendental reflection in the manner of Hegel's own earlier essay, 'Faith and Knowledge'). One might ask, however, whether *all* transcendental arguments, even those that are no longer tied to a specifically Kantian understanding of the world, suffer from the same problems. It is beyond the scope of this chapter to consider this question in any detail. Nonetheless, I would like to suggest here—partly out of conviction, but also partly to provoke further discussion—that such arguments must suffer from the same problems by their very nature. The reason, as I see it, is this: transcendental philosophy aims to reach its own insight into the necessary conditions of, say, the skeptical attitude, an insight that goes beyond what the skeptic himself can see and endorse; such philosophy must, therefore, make some assumption of its own over and above what the skeptic assumes, for if it did not, nothing would take the philosopher beyond what is apparent to the skeptic himself; in that case, however, no specifically *transcendental* insight could be gained by the philosopher. This must be true, I would argue, even in cases in which the transcendental philosopher insists that he begins from nothing but what is endorsed by the skeptic himself: transcendental philosophy by its very nature starts from some assumption of its own and

3. Hegel and Phenomenology

After 1804 Hegel insists that genuine philosophy should be radically self-critical. It may not therefore simply beg the question against ordinary consciousness.³⁰ Ordinary consciousness is perfectly entitled simply to assume that its position is right and philosophy's wrong, but the converse is not true for philosophy. For this reason in its encounter with ordinary consciousness philosophy may not appeal to itself and its own insights and assumptions in order to demonstrate its superiority over the ordinary point of view. It may not rely on such assumptions to argue directly *against* ordinary consciousness; nor may it rely on them to uncover truths in, or conditions of, ordinary consciousness that such consciousness itself cannot see. So what can it do? For Hegel, there is only one alternative: phenomenology, in which it is shown that the commitments of ordinary consciousness *itself* lead beyond the standpoint of ordinary consciousness to that of 'absolute knowing' or speculative philosophy.³¹

If it is to avoid the problems that beset transcendental philosophy, phenomenology may not begin with any assumptions on the part of the philosopher-phenomenologist concerning consciousness, but must start with the most minimal description of consciousness that ordinary consciousness itself can accept. This most minimal description is set out in these lines from the Introduction to the *Phenomenology*:

Consciousness simultaneously *distinguishes* itself from something, and at the same time *relates* itself to it, or, as it is said, this something exists *for* consciousness; and the determinate aspect of this *relating*, or of the *being* of something *for a consciousness*, is *knowing*. But we distinguish this being-for-another from *being-in-itself*; whatever is related to knowledge or knowing is also distinguished from it, and posited as *existing* outside of this relationship; this being-in-itself is called *truth*. (PS 52–3/76)

The language used in these lines may not be particularly ordinary, but the description of consciousness that they contain does not go beyond what consciousness itself would acknowledge. Consciousness distinguishes itself from something: it takes itself to be consciousness *of* something. Moreover, it takes what it is conscious of actually to *be there*: it considers itself to be aware not merely of the 'idea' or 'representation' of a thing, but of the thing itself. The object of consciousness is thus regarded in two ways. On the one hand, the object is held by consciousness to be whatever *it* is and to have an independent character that is not constituted by our consciousness of it; that is, it is held to be 'being-in-itself'. On the other hand, the object as it is in itself is also held to be available to and known by consciousness: when we look at the thing, we do not ordinarily

argues, partly on the basis of that assumption, to some conclusion of its own, for if it did not begin from some such assumption, it would not be able to see any more than the skeptic himself is capable of seeing.

A further question that should perhaps be considered is whether my argument that Hegel's phenomenology is not an essay in transcendental argumentation has any bearing on Husserl's project of 'transcendental phenomenology'. The answer to this question will depend largely on who one thinks has the more rigorous conception of phenomenology: Husserl or Hegel.

³⁰ See Westphal 2000, 288.

³¹ See Houlgate 2005a, 63–6.

suspect that some deep dimension of its being remains hidden from us, but we assume that what the thing is is perfectly clear to us (or at least will become clear if we look hard enough). In that sense, consciousness takes its knowledge of the object to be knowledge of what the object *is*, and so, in Hegel's language, understands knowledge to be the '*being of something for a consciousness*'.

It is very important to recognize that what Hegel takes himself to be giving in these lines is not his own philosophical description or theory of consciousness. He is presenting what he takes to be an uncontroversial description of consciousness that would be endorsed by consciousness itself. This becomes especially clear a little further on in the Introduction. Hegel writes that 'consciousness is, on the one hand, consciousness of the object, and on the other, consciousness of itself'. He then notes that consciousness is in a position to compare the object with its knowledge of the object, 'since both are *for* the same consciousness' (PS 54/77–8). It is consciousness *itself*, therefore—rather than the philosopher—that draws the distinction between consciousness and its objects. The philosopher or phenomenologist offers no ideas of his or her own concerning the constitution of consciousness (or intuition or sensation).

Furthermore, the phenomenologist does not proceed by making other, more detailed claims of his or her own about consciousness. In particular, the role of the phenomenologist is not to determine the 'conditions' or 'presuppositions' of ordinary consciousness that may not be apparent to consciousness itself. The phenomenologist's task is, rather, to trace the *experience* that consciousness itself makes (and must make) of its object. In this respect, phenomenology is very different from transcendental philosophy: for, unlike the latter, phenomenology never becomes a philosophical theory *about* consciousness and its conditions, but remains throughout a study of what emerges (and must emerge) *in* and *for* consciousness itself.³²

This is not to say that the phenomenologist is utterly inactive in the course of phenomenology. Phenomenology is not an empirical discipline that simply describes the myriad ways in which our personal, social, and historical experience actually develops or has developed. It traces the experience that, *logically*, consciousness must make, given the way it initially understands its object. In this sense, Hegel's *Phenomenology* offers a philosophical account of the development of consciousness that has to be articulated by the phenomenologist.³³ Yet what it presents is a philosophical account that starts from consciousness's *own* understanding of its object and shows how various further aspects of that object (and of consciousness) necessarily *become apparent to consciousness itself*. Phenomenology does not argue from features of consciousness which the philosopher takes for granted back to conditions of consciousness that only the philosopher can see (or that consciousness itself can see only with the help of the philosopher).

³² See Wieland 1973, 77: 'something is only then accepted as an insight, if consciousness has come upon it itself' (my translation).

³³ See Houlgate 2005a, 55.

I have argued elsewhere that the transition from one shape of consciousness to another in the *Phenomenology* is actually made by the phenomenologist, rather than the specific shape of consciousness under discussion. Sense-certainty does not itself turn into perception, nor does the slave himself become a stoic. Rather, *we* the phenomenologists make the transition to a new shape of consciousness that takes up explicitly what has emerged in the experience of the preceding shape.³⁴ Once again, therefore, it is clear that the phenomenologist has plenty of work to do in Hegel's new science. This does not, however, invalidate the interpretation put forward here: for the transition that we make to a new shape is made necessary by the *experience* of consciousness itself, not by any particular insight on our part into the conditions of such experience.

Hegel's *Phenomenology* certainly presents the development of consciousness in a way we would not ordinarily present it. It is a logical account of the development of consciousness that employs distinctive philosophical categories, such as 'being-for-self' or 'being-for-another' that in this precise form are not part of our ordinary vocabulary (see *PS* 35, 77/56, 104). In this sense, Hegel's text provides a philosophical understanding of ordinary consciousness. Yet it does not offer distinctively philosophical theses *about* consciousness or its conditions from a perspective that is assumed from the start to be superior to that of ordinary consciousness. Rather, it employs philosophical categories to articulate the development that is made necessary by the *experience of ordinary consciousness itself*. The phenomenologist, therefore, endeavours not to beg the question against ordinary consciousness from the start, but to discover where the commitments of ordinary consciousness itself lead (in the hope that they will lead to the standpoint of philosophy). By contrast, the transcendental philosopher, at least as exemplified by Kant, sets out the conditions that *he* determines to be necessary, given what *he* has assumed intuition and sensation to entail. (This is also true, as I have suggested, of Taylor's Hegel.)

On my interpretation, therefore, Hegel's *Phenomenology* is a radically non-transcendental text. This is important to bear in mind when considering Hegel's remarks about the relation between the consciousness of objects and the consciousness of oneself. As we have seen, Kant claims that the 'I think' must be able to accompany all my representations; that is to say, it must be possible for me to think of my representations *as mine* and so to be conscious of myself in being conscious of objects. Kant also claims that another, more implicit form of self-consciousness constitutes the transcendental condition of experience. This self-consciousness—the synthetic unity of apperception—is not one of which we can ever be directly aware, but it must be assumed or posited if the possibility of experience is to be explained.

In the *Phenomenology* Hegel also connects consciousness and self-consciousness (or self-awareness). His point, however, is somewhat different from Kant's. Hegel

³⁴ See Houlgate 2005a, 56–7, and Hegel, *PS* 55–6/79–80.

claims that ordinary consciousness *itself* draws a distinction between itself and its object: 'consciousness is, on the one hand, consciousness of the object, and on the other, consciousness of itself' (PS 54/77). Such self-consciousness is not, therefore, a transcendental condition of consciousness of objects but rather a constitutive moment of such consciousness itself. It is the consciousness of oneself in relation to which one regards what one sees as that which is *different from me* and thus as an *object* in its own right. According to Hegel, therefore, it is not just the case that it must be possible for consciousness to be self-conscious, but a certain consciousness or awareness of oneself is always part of the experience of consciousness itself. In other words, the 'I think'—or at least the 'I'—always accompanies consciousness to a certain degree.

It is clear even from the opening account of sense-certainty, however, that the self-consciousness or self-awareness that is entailed by consciousness can be more or less explicit. The first form of sense-certainty discussed by Hegel has as its explicit object of attention *this, here, now*, not the 'I'. The second form, however, has as its object of attention what *I* mean by this (PS 61–2/86–7). In contrast to the first form, therefore, the second is more explicitly self-aware. In this sense, one might after all say that, for Hegel, an explicit sense of self must be able to, but need not actually, accompany all my representations. However this may be, what is clear is that, in the *Phenomenology*, Hegel is interested in the process that leads logically to the emergence of more explicit self-consciousness *in experience*. His aim is not to argue, from a superior philosophical position, that self-consciousness is the transcendental condition of consciousness.

It should be noted, by the way, that the 'self-consciousness' Hegel identifies as a moment of consciousness in the chapter on sense-certainty (and also in the following two chapters) is to be distinguished from the self-consciousness that is the focus of attention in chapter 4 of the *Phenomenology*, 'The Truth of Self-Certainty'. The latter is no longer simply one aspect of the consciousness of objects, but has *itself* as the principal object of its concern. Its goal is to achieve freedom for itself and it subordinates everything else to this goal. Indeed, it understands its freedom to require the negation of other things to a greater or lesser degree.³⁵ This latter form of self-consciousness is not itself a moment of consciousness, but rather incorporates consciousness as a moment within itself. It is this fact that places self-consciousness in relation to the objects of consciousness and ultimately explains why self-consciousness initially takes the form of desire (*Begierde*).³⁶

Before we proceed there is one further point to consider. In chapter 6 of the *Phenomenology*—on 'spirit' (*Geist*)—Hegel writes as follows:

Spirit is thus self-supporting, absolute, real being. All previous shapes of consciousness are abstract forms of it. They result from spirit analysing itself, distinguishing its moments, and dwelling for a while with each. This isolating of those moments *presupposes* spirit itself and *subsists* therein; in other words, the isolation exists only in spirit which is a concrete existence. (PS 264/325)

³⁵ See Houlgate 2005a, 68.

³⁶ See Hegel, PS 105/138–9, and Houlgate 2003, 8–29.

These lines suggest that spirit proves to be the ultimate presupposition or 'condition' of consciousness, self-consciousness, and reason; and this in turn suggests that Hegel's *Phenomenology* does, after all, provide a transcendental account of the conditions of possibility of consciousness and its objects. There is, however, another way of understanding Hegel's statement. Spirit does, indeed, prove to be the 'condition' of consciousness in the sense that consciousness, self-consciousness, and reason prove to be mere moments or aspects of spirit. Yet what shows them to be such moments is ultimately *their own experience*—experience that is initially generated by what sense-certainty takes its object to be, not by any assumptions about consciousness made by the philosopher. Furthermore, consciousness, self-consciousness, and reason prove to be moments of spirit *for spirit itself* (more or less explicitly), not just for the philosopher undertaking the phenomenological analysis. My claim that Hegel's *Phenomenology* is not an essay in transcendental argument is thus not meant to suggest that nothing at all is said in that text about the presuppositions or conditions of consciousness and its objects. It suggests only that whatever understanding of the presuppositions of consciousness emerges does so in and through the experience of consciousness itself, not through the privileged insight (and due to the assumptions) of the philosopher.

It should be noted, however, that although spirit as such or as a whole may well be the presupposition of the shapes that precede it, not every individual shape of consciousness, self-consciousness, reason, and spirit can be understood in this way to be the *presupposition* of the ones that precede it. Stoicism, for example, is not the presupposition of the master/slave relation, and absolute freedom (realized in the French Revolution) is obviously not the presupposition of 'the true spirit' (which is realized in the world of Sophocles' *Antigone* and *Oedipus Rex*). The principal and enduring relation between the shapes discussed in the *Phenomenology* is thus not that each shape is the presupposition of the preceding one; it is, rather, that each one takes up and affirms the *truth*—the true character of the specific object of consciousness or of the specific mode of consciousness itself—that emerges in the experience of the previous shape (see PS 55/78–9).

4. Sense-Certainty

In order to clarify further the difference between Hegel's phenomenological procedure and transcendental argumentation I will now look briefly at the end of Hegel's account of sense-certainty and the transition to perception.

Robert Pippin understands this transition to involve a broadly transcendental move. That is to say, the transition is made from sense-certainty to perception because concepts and universals—the 'properties' of which perception is conscious—are deemed *by the philosopher* to be the presuppositions and conditions of 'even such a minimal apprehension' of objects as sense-certainty. As Pippin puts it, Hegel's goal in the opening chapters of the *Phenomenology* is to 'demonstrate that even the simplest form of demonstrative

reference would not be possible without some describing capacity, a capacity that requires descriptive terms or predicates . . . , not merely deictic expressions and atomic objects.³⁷ It is clear from Pippin's account that sense-certainty itself does not come to an understanding of its own presuppositions; nor, indeed, does the philosopher's understanding of those presuppositions emerge directly out of the experience of sense-certainty. What is shown by the experience of sense-certainty is simply the latter's *failure* to bring to mind objects as it defines them, its failure to 'immediately "mean" its objects'.³⁸ Once that failure has been brought to light, it is a separate task to 'investigate what capacities a conscious subject must be assumed to have in order to experience successfully'.³⁹ This investigation (in the following chapters of the *Phenomenology*) rests, however, on assumptions that Hegel makes from the start of his phenomenological analysis, namely that 'consciousness is clearly a kind of intentionality'—that is, consciousness of a *determinate* object—and 'that intentionality is always a result of a subject's *activity*'.⁴⁰ These assumptions are themselves derived from the way in which natural consciousness understands itself—from the fact that consciousness 'always involves both a "relation to" and a "distinguishing from" an object'⁴¹—but they are not shared by *sense-certainty*, since the latter does not initially take its object to be determinate or itself to be active. The presuppositions of sense-certainty (and of all genuine cognition) are thus worked out by the philosopher on the basis of assumptions that sense-certainty itself does not share and that do not arise out of its experience. The analysis given by Pippin's Hegel of the failure of sense-certainty may well be based upon the experience of objects 'defined as sense-certainty defines them' (and in that sense be an immanent analysis), but what takes us beyond the perspective of sense-certainty to a deeper understanding of the perceptual and conceptual presuppositions of cognition are ultimately assumptions that are alien to sense-certainty. As Pippin himself admits, 'it is because Hegel holds fast to the natural consciousness assumption that consciousness of *determinate* objects is possible, that such discrimination must be possible if there is to be experience at all, that he maintains that this position of sense-certainty, as now revealed in its "experience", must be *overcome*'.⁴² In this respect, therefore, Pippin's Hegel provides a recognizably transcendental account of the move from sense-certainty to perception.

On my reading, by contrast, the phenomenologist is meant to avoid making any assumptions about cognition or about the conditions and presuppositions of consciousness and is meant simply to trace the *experience* that is made by consciousness itself. On this reading of the *Phenomenology* the transition is made to perception not because perception is judged by the phenomenologist to be the condition of determinate experience, but because perception takes up explicitly what emerges in the experience of sense-certainty itself. Perception is made necessary *by sense-certainty alone*, therefore, not by the assumptions and commitments of the phenomenologist.

Let us see in more detail how this transition is made. At the beginning of the third stage of sense-certainty, consciousness tries to bring the immediate object to mind by

³⁷ Pippin 1989, 117.³⁸ Pippin 1989, 120, 122.³⁹ Pippin 1989, 122.⁴⁰ Pippin 1989, 116, my emphasis.⁴¹ Pippin 1989, 116.⁴² Pippin 1989, 120.

pointing that object out to itself and to us. Hegel argues, however, that in the process of pointing out its object, consciousness learns that its object is not in fact the simple, immediate *this* that it took it to be, but rather something complex.

This is especially evident in the consciousness of what I take to be before me *now*, or of what Hegel calls 'the now' (*das Jetzt*). The problem, according to Hegel, is that, in the very act of pointing out the 'now', consciousness moves on from that now to another 'now'. Thus, precisely because it has been *pointed out*—and the act of pointing out takes time for consciousness itself—any such now is necessarily brought to mind at a *later* moment. This means that it can be brought to mind only as a now that no longer simply *is*, but that *has been*. This does not mean that consciousness fails to point out the now it has in mind. In the process Hegel is analysing, consciousness succeeds in its endeavour: the now is, indeed, pointed out and as a result *is now* present to consciousness. It is present, however, as a now that *has been* pointed out. In other words, it is present as a now that necessarily has a past. In the very act that is meant to secure for consciousness its certainty of the *now*, the now thus proves *not* to be the pure immediate presence—the pure and simple *now*—that consciousness has taken it to be. Rather, the now necessarily shows itself, in the very experience of it, to be a now-that-has-been or a continuity of present *and* past moments. As such, the object of sense-certainty proves to be 'an absolute plurality of Nows' (PS 63–4/88–9). Such a complex continuity of different moments is called by Hegel a 'universal'. It is, however, a concrete, internally differentiated universal, rather than the empty, abstract universal that sense-certainty first experienced the *this* and the *I* to be.

The movement of pointing out is thus 'the experience of learning that Now is a *universal*' (PS 64/89). Nonetheless, Hegel notes, sense-certainty still 'wants to apprehend the *This*' (PS 67/93). In spite of all it has learned about its object, therefore, sense-certainty does not relinquish its identity as *sense-certainty* and cease seeking immediate certainty of *this*, *here*, *now*. Sense-certainty does not, therefore, grasp the full implication of its own experience and accept that the object of consciousness has actually been shown to have a new logical form.

The shape of consciousness that does accept this and 'take over the truth' revealed by sense-certainty is perception or *Wahrnehmung* ('taking truly'). 'Perception', Hegel tells us, thus 'takes what is present to it *as a universal*' (PS 67/93, my emphasis). In Hegel's account, sense-certainty does not turn itself around and become perception; it is *we* who make the transition to perception. Nevertheless, this transition is made necessary by the experience of sense-certainty itself, since perception is simply the shape of consciousness that *accepts* the truth disclosed by sense-certainty and takes the object as it is actually known by sense-certainty—the 'being-*for*-consciousness of the first in-itself'—to be the *new* object of consciousness (see PS 55/79). In the transition that *we* make, consciousness itself thus mutates immanently and logically—through its own experience—into a new shape.

Note that Hegel's account of sense-certainty is not a transcendental one. Like Kant, Hegel points out that the consciousness of what is present is inseparable from the consciousness of what is past. But he does not explore the transcendental conditions that, in

Kant's view, make such consciousness possible, namely the synthesis of reproduction and transcendental apperception. Indeed, Hegel evinces no interest in explaining how *from his own perspective* sense-certainty is able to hold past and present moments together: he says nothing, for example, about the role that memory or attention might play in making sense-certainty possible. His concern is simply to show that a consciousness of what is past necessarily emerges *in the experience of sense-certainty itself*, through the very act of pointing that is meant to secure for consciousness pure presence.

In so doing Hegel does not (like Kant) make any assumptions of his own about time: he does not presuppose that time is divided up into atomic units called 'nows'. His claim, rather, is that sense-certainty itself conceives of its object as the simple *now*, but necessarily goes beyond such simple immediacy (and realizes that it does so) in the very act of pointing it out to itself. Hegel's argument is thus rigorously phenomenological.

The transition to perception is not to be understood as a transcendental one either. Perception is not introduced by the phenomenologist as the condition of determinate experience. Rather, perception simply takes up what emerges in the *experience of sense-certainty itself*. This pattern is repeated throughout the *Phenomenology*. New shapes of consciousness are not introduced because they are judged by the phenomenologist to constitute more concrete conditions of experience but arise because they take up and render explicit what is revealed in the experience of the shapes that precede them. This is true of every shape of consciousness (in the broad sense) after sense-certainty, whether it belongs to consciousness *sensu stricto*, to self-consciousness, to reason, or to spirit. At no point does (or, at least, should) the phenomenologist intervene to make claims of his own about the conditions of experience; the whole development is driven forward by the experience of consciousness itself. As Lauer puts it, 'in the consistently sustained *experience* of the object the object reveals itself more and more as what it truly is'.⁴³ At the end of the *Phenomenology*, therefore, we do not reach what the *phenomenologist* considers to be the ultimate condition of experience. Rather we discover what is finally revealed in the *experience* of consciousness itself, namely the identity of thought and being. (As I have noted, this does not mean that nothing at all is disclosed in the *Phenomenology* about the presuppositions of consciousness. It means simply that such presuppositions are revealed through the experience of consciousness *sensu stricto*, self-consciousness and reason themselves and come to be known—more or less explicitly—by spirit itself.)

Pippin interprets Hegel as a broadly transcendental philosopher. Accordingly, he understands the *Phenomenology* to culminate in a transcendental claim made by the philosopher: namely, that 'the forms of the subject-object, or Notion-Truth opposition inherent in natural consciousness . . . themselves presuppose the speculative understanding of a subject-object or Notion-Truth identity'.⁴⁴ The task of Hegel's *Logic*, for Pippin, is to articulate the basic structure of that speculative understanding. The categories set out in the *Logic* do not, therefore, grant us direct insight into the nature of being—they do not show us what there *is*—but they set out the ultimate conditions of determinate cognition

⁴³ Lauer 1976, 37.

⁴⁴ Pippin 1989, 112.

(and of the objects of cognition). In Pippin's own words, the *Logic* does not demonstrate that 'being as it is in itself can be known to be as we conceive it to be', but it sets out 'all that "being" could intelligibly be' for reflective subjectivity.⁴⁵ Pippin's transcendental approach to the *Phenomenology* thus leads him directly to a non-ontological understanding of the *Logic*.⁴⁶

On my reading, by contrast, the *Phenomenology* discloses the truth that emerges in the experience of consciousness itself. Consciousness starts out claiming to be consciousness of *being*, of what there *is*. As a phenomenologist, Hegel must accept this claim and may not challenge it with counter-claims of his own. Furthermore, his task is not to point to the conditions that would make such consciousness possible. It is simply to show the various forms that such being takes in the experience of consciousness itself (for example, that of a mere object, of reason, of spirit). At the close of the *Phenomenology*, in the chapter on 'absolute knowing', being shows itself in experience to have the structure of the self-determining 'concept' (*Begriff*) (PS 491/589).⁴⁷ The task of the *Logic* is thus not to lay down the conditions of determinate cognition, but to articulate in detail the conceptual structure of *being itself*. The *Logic* is not a work of transcendental philosophy, therefore, but a work of full-blown *ontology*.⁴⁸

Note, however, that if speculative logic is to be a genuinely self-critical discipline it may not begin immediately with the concrete conception of being-as-concept that emerges at the end of the *Phenomenology*. The fully self-critical speculative logician must suspend all inherited conceptions of being—including that inherited from the *Phenomenology*—and start from scratch by considering the pure, indeterminate thought of being as such.⁴⁹ Nonetheless, the *Logic* takes off from the *Phenomenology* since it determines anew the conceptual structure of being that emerges in the *experience* of consciousness. *Pace* Pippin, the *Logic* does not continue the (allegedly) transcendental project of the *Phenomenology*.

This, however, raises one last question: if Hegel's *Logic* is an ontology and not a form of transcendental philosophy, how can Hegel be considered to be a genuinely *post-Kantian* thinker? The answer is simple: for Hegel the imperative for all post-Kantian philosophy is not, as Pippin maintains, to avoid direct claims about being-in-itself and to restrict oneself to determining the 'conditions of the possibility of knowledge'.⁵⁰ It is to be radically self-critical and to avoid arbitrary assumptions.⁵¹ This requires of us, especially, that we avoid begging the question against ordinary consciousness in the manner of transcendental philosophy by pointing to conditions of consciousness that such consciousness cannot see if left to its own devices. For Hegel,

⁴⁵ Pippin 1989, 98, 115.

⁴⁶ Pippin himself would reject the idea that there is actually a distinction between transcendental logic, as his Hegel conceives it, and ontology, since such logic, and the self-knowledge that it involves, 'counts as the knowledge of the thing in itself that Kant paradoxically denied' and so counts as ontology or metaphysics (see Pippin 1989, 98, 180). Nonetheless, Pippin's denial that, for Hegel, thought can make any 'direct (i.e., not transcendently deduced) claim about reality' (Pippin 1989, 190) prevents Hegel's logic from being an ontology in the strong sense in which Spinoza and, in my view, Hegel himself would understand it.

⁴⁷ See also Houlgate 2005a, 63–6.

⁴⁸ See Houlgate 2006, 115–43.

⁴⁹ See Houlgate 2005a, 65; 2006, 162.

⁵⁰ Pippin 1989, 95, 98.

⁵¹ See Houlgate 2006, 24–8.

therefore, the appropriate discipline for philosophy to pursue after Kant is either presuppositionless ontological logic or phenomenology as the non-transcendental account of the experience of consciousness. There is no other option.

References

- Allison, Henry (2004), *Kant's Transcendental Idealism: An Interpretation and Defense*, revised and enlarged edn (New Haven: Yale University Press).
- Hegel, G. W. F. (1970a), 'Glauben und Wissen', in *Jenaer Schriften (1801–7)*, ed. E. Moldenhauer and K. M. Michel, *Werke in zwanzig Bänden*, ii (Frankfurt am Main: Suhrkamp Verlag), 287–433.
- Hegel, G. W. F. (1970b), *Phänomenologie des Geistes*, ed. E. Moldenhauer and K. M. Michel, *Werke in zwanzig Bänden*, iii (Frankfurt am Main: Suhrkamp Verlag).
- Hegel, G. W. F. (1977), *Phenomenology of Spirit*, tr. A. V. Miller (Oxford: Oxford University Press). [PS]
- Hegel, G. W. F. (1999), *Science of Logic*, tr. A. V. Miller (Amherst, NY: Humanity Books).
- Houlgate, Stephen (2003), 'G. W. F. Hegel: *The Phenomenology of Spirit*', in Robert Solomon and David Sherman (eds), *The Blackwell Guide to Continental Philosophy* (Oxford: Blackwell), 8–29.
- Houlgate, Stephen (2005a), *An Introduction to Hegel: Freedom, Truth and History* (1991; Oxford: Blackwell).
- Houlgate, Stephen (2005b), 'Glauben und Wissen: Hegels Immanente Kritik der Kantischen Philosophie oder die (Illegitime) "Ahnung eines Besseren"', *Hegel-Jahrbuch*, 3, 152–8.
- Houlgate, Stephen (2006), *The Opening of Hegel's Logic: From Being to Infinity* (West Lafayette, IN: Purdue University Press).
- Kant, Immanuel (1997), *Critique of Pure Reason*, tr. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press). [CPR]
- Lauer, Quentin (1976), *A Reading of Hegel's Phenomenology of Spirit* (New York: Fordham University Press).
- Longuenesse, Béatrice (1998), *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*, tr. Charles T. Wolfe (Princeton: Princeton University Press).
- Pippin, Robert (1989), *Hegel's Idealism: The Satisfactions of Self-Consciousness* (Cambridge: Cambridge University Press).
- Pippin, Robert (1993), 'You Can't Get There From Here: Transition Problems in Hegel's *Phenomenology of Spirit*', in Frederick Beiser (ed.), *The Cambridge Companion to Hegel* (Cambridge: Cambridge University Press), 52–85.
- Taylor, Charles (1976), 'The Opening Arguments of the *Phenomenology*', in Alasdair MacIntyre (ed.), *Hegel: A Collection of Critical Essays* (1st publ. 1972; Notre Dame, IN: University of Notre Dame Press), 151–87.
- Westphal, Kenneth (2000), 'Kant, Hegel, and the Fate of "the" Intuitive Intellect', in Sally Sedgwick (ed.), *The Reception of Kant's Critical Philosophy: Fichte, Schelling, and Hegel* (Cambridge: Cambridge University Press), 283–305.
- Wieland, Wolfgang (1973), 'Hegels Dialektik der sinnlichen Gewissheit', in Hans Friedrich Fulda and Dieter Henrich (eds), *Materialien zu Hegels 'Phänomenologie des Geistes'* (Frankfurt am Main: Suhrkamp Verlag), 67–82.

9

Transcendental Aspects, Ontological Commitments, and Naturalistic Elements in Nietzsche's Thought

Béatrice Han-Pile

I

In what follows I shall focus on Nietzsche's views on knowledge and leave aside such questions as to whether, for example, his understanding of morality should be considered as naturalistic or not. Even from this restricted perspective, the interesting thing about Nietzsche is that he is a particularly tricky case: he has been characterized as a philosopher in the transcendental tradition (Green 2002) and more often as a naturalist and an empiricist (Leiter 2002; Clark 1990, 2001; Richardson 2004; and in a more nuanced way, Cox 1999).¹ Prima facie (and perhaps worryingly), there are justifications for both views, in spite of their apparent incompatibility. Thus, Nietzsche has claimed that human experience is necessarily structured according to '*a priori* forms' (TL 87). He has argued for a form of transcendental idealism, making both the ontological assumption that there is a way in which things are in themselves and the epistemological claim that such things are a 'mysterious X' (TL 83), by definition beyond the scope of human knowledge. Yet in a more naturalistic fashion he has also suggested that philosophy should model its inquiry and method on the sciences, and take

¹ Contrary to Clark and Leiter, Cox takes into account Nietzsche's repeated attacks against science: thus Nietzsche 'uncover[s] a residual theology in the modern scientific project's claim to describe the way the world really is' (Cox 1999, 6). Cox argues that science must 'overcome itself' into aesthetic discourse, and that such self-overcoming is 'more rigorously naturalistic' (probably in the sense that it is more thoroughly non-metaphysical) than the scientific project—a view which presumably neither Clark nor Leiter would be happy to endorse.

their results into account. In spite of his disagreement with Darwin on the question of the preservation of life, he has promoted a view of human nature as determined by evolutionary principles. Furthermore, he has rejected both the idea of a *Hinterwelt* and of the appearances/thing in itself distinction, in a way which seems to preclude the dualism inherent in transcendental idealism. And as if endorsing central elements of both naturalism and transcendental philosophy was not tricky enough, Nietzsche also seemingly held a view that denies a central feature shared by these two trends: he apparently rejected the possibility of objective knowledge (his so-called error theory). This has led others (in particular Derrida 1979; De Man 1979; Kofman 1993) to think that he is a proto-deconstructionist, whose purpose is to undermine the ideal of secure knowledge which underlies both naturalistic and transcendental projects.

The main way to at least reduce the confusion is to look at things chronologically. In this regard, the dominant interpretation (put forward by pro-naturalistic interpreters such as Maudemarie Clark and Brian Leiter) is that while the early Nietzsche, under the influence of Kant and Schopenhauer, was a transcendental idealist, he soon renounced the mistaken ways of his youth and became a naturalist. To put it simply, the Nietzsche of 'Truth and Lies' believed both in the existence of things in themselves beyond the empirical realm and in their unknowability. Correlatively, this early commitment to transcendental idealism is seen as the main ground for his error theory: since its empirical nature prevents it from capturing the essence of things in themselves, all human knowledge is by definition erroneous. Yet from *Human, All Too Human* onwards Nietzsche would have started to have doubts about the existence of things in themselves, and ultimately rejected the notion as well as the rather inconvenient error theory it warranted. This would have paved the way for a naturalistic epistemology characterized by his alleged enthusiasm for the sciences,² his emphasis on the continuity between philosophy and science, and a 'non-metaphysical' commitment to what Clark calls 'common sense realism', i.e. the belief in the existence of mind-independent empirical objects (Clark 1990, 61; see ch. 2 in general).

In my view, the problem with this neat reconstruction is that it oversimplifies both Nietzsche's early position and his philosophical development.³ The main reason for

² There are many other passages, both in the published and unpublished work, which suggest (contra Clark or Leiter) that science does not enjoy any particular epistemic privilege for Nietzsche: thus 'science at its best seeks most to keep us in this simplified, thoroughly artificial, suitably constructed and suitably falsified world . . . —it loves error because, being alive, it loves life' (*BGE* ii, §24). Or again, 'physicists believe in a "true world" of their own fashion: a firm systematisation of atoms in necessary motion. . . . But they are in error. The atom they posit is inferred according to the logic of the perspectivism of consciousness' (*WP*, §636). Perhaps the most lapidary formula is the following one: 'ultimately, man finds in things nothing but what he himself has imported into them; the finding is called science' (*WP*, §606).

³ Another difficulty, noted by many commentators (in particular Poellner, Cox, and Green) is that this reconstruction is not supported by uncontroversial textual evidence, and leaves out the entirety of the *Nachlass*. Having said that, I should say that, in other regards, I am sympathetic to Leiter and Clark's approach, which I have often found illuminating. In particular, I completely agree with Leiter's criticism of the interpretations of Nietzsche that see him as a precursor of postmodernism intent on denying the possibility of any truth and objectivity (see Leiter 1994).

this is that it implicitly rests on two unwarranted assumptions: first, that transcendental idealism can be uncontroversially identified with a two-world view and the strong ontological commitment to the autonomous existence of things in themselves it entails; secondly, that being a transcendental idealist in this robust sense is the only way to cash out transcendental claims.⁴ Yet as is well-known, deflationary interpretations such as Bird's (Bird 1962) or Allison's (Allison 1983) reject the two-world view in favour of a two-aspect reading which radically minimizes the ontological commitments of transcendental idealism in order to emphasize its epistemological ambit, i.e. its attempt to secure the possibility of objective empirical knowledge by identifying its a priori necessary conditions. Interpretations such as Strawson's (Strawson 1966) go even further in this direction by arguing that the best way to bolster this epistemological core is to strip away the ontological commitments entirely. In this chapter, I shall suggest that, likewise, Nietzsche had a much more nuanced view of transcendental idealism than is attributed to him by his pro-naturalistic readers, and that he too was concerned with prying apart its ontological and epistemological dimensions. By means of a close re-examination of a text central to Clark's own interpretation, namely 'Truth and Lies' (henceforth *TL*), I shall establish that his early views about the existence of things in themselves are accompanied by a separate analysis of perception which concludes that our experience now has what looks like transcendental conditions⁵ in that it is necessarily structured

⁴ There are other, related problems with the naturalistic line of interpretation. Of particular relevance is the fact that it rests on a mistaken definition of transcendental philosophy as 'the search for a priori modes of knowledge which might allow knowledge of the thing in itself' (Clark 1990, 116)—along the same lines, another passage defines metaphysical or a priori knowledge as 'non-empirical access to reality and a basis for rejecting the relevance of sense testimony' (Clark 1990, 106). Yet Kant's ambition was not to reject the relevance of sense testimony, but to show that it is not sufficient, per se, to account for the possibility of knowledge. Moreover, defining transcendental knowledge in this way is tantamount to reverting to a pre-Kantian understanding of the a priori: for Descartes or Leibniz, having a priori knowledge is indeed equivalent to having innate ideas which provide us with non-empirically acquired information about such metaphysical issues as the nature of the soul or the existence of God. But as is well known, Kant was as keen as Hume (although for different reasons) to reject the idea that we could have a priori knowledge of metaphysical essences. More generally, many of the naturalistically inclined Nietzsche interpreters often do not define what they mean by 'transcendental philosophy', nor the ways in which they think it is opposed to naturalism. As a result, the criteria they use to argue in favour of Nietzsche's 'naturalism' are often indecisive, because they would equally well apply to a transcendental philosopher and/or rely on an unsuitable view of transcendental philosophy. The reasons for this, I think, are partly historical, and partly due to the context of Nietzsche studies. Thus the rise of naturalism (and of proto-naturalist movements such as materialism in the eighteenth century or positivism in the nineteenth century) was mostly targeted at *metaphysics* rather than at transcendental philosophy. There are traces of this e.g. in the often quoted Quinean definition of naturalism as the rejection of 'first philosophy'. Consequently many of the arguments offered in favour of the naturalistic reading of Nietzsche are grounded in his rejection of metaphysics. Correlatively, this anti-metaphysical bend is accentuated by the perceived need to oppose interpretations—one of the first and most famous being Heidegger's—which emphasize the resurgence of metaphysical aspects in Nietzsche's thought. However as I have indicated, Kant was equally distrustful of what he called 'dogmatic' metaphysics. Because they often do not take this into account, the naturalistically inclined readers of Nietzsche, while they have little trouble showing that he does not want to be a metaphysician, fail to argue convincingly against his being a transcendental philosopher.

⁵ By this I mean, in the most general way possible, conditions which (a) are non-empirical, non-causal conditions of possibility; (b) are modally necessary and presuppose a normative conception of knowledge, prescribing what must be the case (rather than just describing what is the case) in order for us to be able to

by 'a priori forms' (TL 87), amongst which time, space, succession, and coexistence. These conditions are identified independently from any ontological commitments, through a proto-genealogical study of the evolution of human perception. Given this dual angle, it is no longer possible to conclude, as Leiter and Clark do, that Nietzsche's later revision of his early views about things in themselves is tantamount to a rejection of transcendental idealism *simpliciter*, nor that it results in the elimination of all possible transcendental aspects in his thought.⁶ In fact, his reflection on his previous ontological commitments is much more subtle than is usually allowed for: as we shall see, it warrants the passage from a robust to an ultra-deflationary position which establishes the need to remain agnostic about ontological commitments (rather than rejecting them, which per se is another form of robust commitment). Correlatively, Nietzsche's philosophical itinerary can be fruitfully interpreted as a deepening of his early thoughts on the nature of the conditions of possibility of experience, and thus as the development of the epistemological heritage of transcendental idealism.

In a similar way, the so-called error theory is not so easily explained away by Nietzsche's abandonment of his earlier beliefs in the existence of things in themselves. As seen by Clark, it does remove one of its grounds, namely the impossible demand for a metaphysical adequation between human knowledge and the essence of things considered in themselves. However, TL offers a second, independent argument for the error theory: regardless of any ontological commitments, it infers from the contrast between our current perceptual experience and a different, more primordial form of experience (relegated to a mythical past) that the structuration of the former by a priori elements is *in itself* a form of falsification. As we shall see, this early claim is developed by the later Nietzsche's reflection on the perspectival conditions of life and their relation both to the primal stream of our impressions and what he calls the 'world of becoming'. Unless an alternative interpretation which does not simply rest on Nietzsche's rejection of his early robust ontological commitments is proposed, the error theory will thus remain a serious obstacle to all interpreters (naturalistically *and* transcendently inclined) who seek to find in Nietzsche's thought a ground for the possibility of objective knowledge.

Once the complexity of the relation between ontological commitments, transcendental aspects, and naturalistic elements in Nietzsche's work is fully acknowledged, new questions arise, on which I shall focus in the second part of this chapter: clearly the thrust of Nietzsche's analysis in TL is *both* to assert the existence of 'transcendental' conditions *and* to show that they have an empirical genesis, which of course throws

know anything: (c) consequently, involve a bid for universality and are governed by a distinctively foundationalist ambition (grounding and thus securing against skepticism the possibility of objective knowledge). I shall discuss the notion further along in the chapter, in particular when examining Nietzsche's naturalization of the transcendental.

⁶ Cf. Clark 1990, 61: 'in fact there is nothing at all transcendental about the position I attribute to Nietzsche. Nothing I attribute to him is out of accord with Rorty's view that "philosophy will have no more to offer than common sense (supplemented by biology, history, etc.) about knowledge and truth".'

their transcendental status into question (hence my use of quotation marks) and attests to the presence of strong naturalistic elements in his thought (in particular because of his Humean insistence on the part played by belief, habit, and social practices in the genesis of our 'a priori' forms). So how should we construe Nietzsche's so-called 'naturalization' of the transcendental,⁷ and what are its consequences for the idea of transcendental conditions (in particular, regarding their modality and their scope)? Secondly, how do the ontological and epistemological dimensions of transcendental idealism fare in later Nietzsche's work? Finally, and most importantly, is it possible to interpret his error theory in a way that doesn't radically threaten the possibility of his having coherent views about knowledge (be they naturalistic or transcendental)?

Allow me one more remark before I turn to a close examination of these issues: it should be clear from what precedes that my aim in this chapter will not be to make a case for Nietzsche being either a naturalist or a transcendental philosopher. I do not think that it is possible to decide either way, let alone to identify, as Clark and Leiter do, an evolution from an early, transcendentially inclined position to a more mature, naturalistic one. For reasons that will appear progressively, I don't even think that one *should* decide either way. The idea here is rather to examine the interaction of naturalistic and transcendental elements in Nietzsche's thought with a view to identifying nodes of tension, and the way they unfold in his work. As we shall see, such tensions focus mostly on (a) the question of whether experience is actively constituted by the mind or passively received through the senses, (b) the status of the constituting forms and concepts (a priori or not), (c) the problem of whether such a conception of experience requires the existence of a mind-independent reality (and if so, how experience relates to the latter), and (d) depending on the answers provided to the first three questions, the scope and validity of human knowledge. In the course of examining these tensions, I shall suggest that perhaps the most profitable hermeneutic hypothesis is to see Nietzsche's struggle with them as an attempt to overcome the strict opposition between naturalism and transcendental philosophy. It is this third way, with its theoretical benefits and costs, which I want to explore in this chapter.

II

What makes *TL* particularly interesting is that it can be seen as a matrix in which most of the strands of Nietzsche's views on knowledge are already present. The text is notoriously complex, and has been the object of numerous interpretations.⁸ While it is obviously an attack against the possibility of universal truth and objective knowledge, what

⁷ The expression was coined by George Stack. However, he was mostly interested in showing the extent of Alfred Lange's influence on Nietzsche's thought; my own concern is not so much with the history of ideas as with the theoretical implications of such a move, both for transcendental philosophy and naturalism.

⁸ In particular Paul De Man, Sarah Kofman, and Clark. While De Man and Kofman read *TL* as a deconstructionist analysis of language (as metaphorical), Clark holds that the main object of the text is to provide a Schopenhauerian criticism of representation as fundamentally inadequate to the in-itself. While I agree

is far less clear is Nietzsche's strategy. The main reason for such complication is that *TL* constantly intertwines two main lines of argumentation, which both have their source in a reflection on transcendental idealism. On the one hand, and most obviously, Nietzsche infers from the lack of metaphysical correspondence between empirical knowledge and things in themselves that our knowledge is 'anthropomorphic' and therefore invalid from a transcendental point of view. As we shall see, *per se* this is not a particularly good point. Yet there is a more subtle argument: Nietzsche is concerned, not only with what he sees as the failure of human knowledge to adequately capture the in-itself, but also with describing the conditions of possibility of experience in a way that remains agnostic about their possible ontological implications. In doing so, he first presents an anti-rationalist view of perception as 'metaphorical': such a view can be interpreted as *both* anti-Kantian (it rejects the claim that experience is constituted by means of objective judgements) and anti-empiricist (it denies that there are any raw sense data). It then turns out, however, that this view is supposed to describe the way things were experienced in a mythical and forgotten past, and does not reflect the way *we* perceive anymore. In order to explain the changes in perceptual conditions, Nietzsche offers a fairly Humean description of the ways in which conceptualization was made necessary by the development of new social practices (the need for truthfulness) and habits. This proto-genealogical account, however, culminates unexpectedly in a quasi-Kantian vision of experience as now shaped by '*a priori* forms' (*TL* 87). Before analysing the substantial changes the second account brings to the idea of transcendental conditions, I shall endeavour to flesh out and clarify the successive steps of this rather dizzying argument, with a view to bringing out both its transcendental and naturalistic aspects, and their mutual tensions.

As seen by Clark, *TL*'s attack against objectivity does not rest on an analysis of language, but of perception (thus the metaphors Nietzsche describes are 'perceptual' (*TL* 84) and, as we shall see, both pre-verbal and pre-linguistic). The perceptual process is detailed as follows: 'a nerve stimulus is transferred into an image: first metaphor. The image, in turn, is imitated into a sound: second metaphor' (*TL* 81).⁹ *Per se*, there is nothing here that suggests an endorsement of transcendental idealism. *Prima facie*, the description is pitched at the empirical level and concerns the genesis of representations. As it is very compressed, it is worth looking at in more detail. According to Aristotle (*Poetics* 1457b), a metaphor is the 'application of a strange term either transferred from the genus and applied to the species, or from the species and applied to the genus, or from one species to another or else by analogy'.¹⁰ No rules or conditions are specified for such transfer (*meta-phorein*). In the case of a transfer from species to

with Clark that the text should be read as an engagement with transcendental idealism, I do not think that this is its only dimension (cf. main text).

⁹ The process is detailed in almost similar terms in a later passage: 'first *images*—to explain how images arise in the spirit. Then *words*, applied to images. Finally *concepts*, possible only when there are words—the collecting of many images in something non-visible but audible (word)' (*WP*, §506).

¹⁰ Aristotle, *Poetics*, tr. H. Fyfe (Cambridge, MA: Harvard University Press, 1973).

genus (or vice versa) there is clearly a relation of kinship between the two terms considered, but this is not the case when the metaphor links two species 'or else'. By virtue of their analogical nature, metaphors point out a similarity between the two terms considered (although strictly speaking an analogy involves four terms, paired two by two, and concerns the relation between the two pairs rather than the terms themselves). But such similarity is not grounded in an objective consideration of the nature of the objects linked: the connection is subjective, and the resulting judgement, taken literally, would be false. The purpose of such a connection is not to form any knowledge of the objects, but to stimulate the imagination so that it bridges the gap between two terms which per se have little in common, and engages in a free play of significations. Thus Paul Eluard's verse, 'the earth is blue like an orange', does not disclose any objective facts about the earth or oranges. However the metaphor brings to mind aspects in which they could be thought of as similar (such as roundness, natural character, fertility, pitted surface, etc.); the way in which it does this is by sparking an imaginative, almost immediate vision rather than by making us reflect on the nature of the objects (although it is possible, as I have started to do, to articulate after the fact some of the connections established).

How does this apply to perception? Nietzsche's idea seems to be that, just as metaphors involve a transfer of meaning from one term to another which does not reflect their objective characteristics so much as the workings of the imagination, in the same way in the case of perception there is no objective, necessary, or rational connection between nerve stimuli, images, and sounds. They are all different in nature and the passage from one to another is only made possible by the subjective power of our imagination, seen as a primal, artistic power upon which there are, originally, no conceptual constraints.¹¹ It has 'no need for [concepts], those makeshifts of indigence' (*TL* 90). Thus we are 'artistically creative subjects' (*TL* 86) and sounds are the 'coagulation of a mass of images which originally streamed from the primal faculty of imagination like a fiery liquid' (*TL* 86). Along the same lines, Nietzsche speaks of the 'drive towards the formation of metaphors' as 'the fundamental human drive' (*TL* 88): we are animated by 'an ardent desire to refashion the world . . . so that it will be as colourful, . . . charming, and eternally new as the world of dreams' (*TL* 89). As it depends purely on the idiosyncrasies of the perceiving individual, each 'perceptual metaphor' is singular and 'without equal'. This also applies to the 'words' used to convey the images: their assimilation to mere sounds, and the fact that they are distinct from concepts (and do not require the latter for their formation) suggests that this proto-language is extremely rudimentary. It clearly has no grammar, no logical connectors, no verbs, etc. Each 'word' is likely to be an onomatopoeia functioning like a proper name, associated by further metaphorical transfer to an individual image after the latter has been independently formed by the imagination. Because there are no conceptual constraints on the

¹¹ Later in the text, Nietzsche will offer a proto-genealogy of the empirical reasons why this primal creative power has become constrained, in particular by the need to find resemblances.

power of the imagination, neither the images nor the words coalesce into an ordered world of representations, but are 'irregular, lacking in results and coherence' (*TL* 89). Correlatively, as there are no common structures for their formation and all perceptual metaphors are private, there is no intersubjectively shared world.

Although it is problematic,¹² such a view is interesting in that it seems aimed at *both* transcendental and empiricist accounts of experience. It is clearly an attack against the Kantian understanding of experience as dependent on judgements which would bind intuitions into objective representations. For one thing, the formation of images as described here is said not to involve any concepts (as we shall see, for Nietzsche the latter arise from the equalization of pre-existing perceptual differences), which removes the possibility of judgement as unification of the manifold under a rule. Correlatively, the idea that representations are perceptual metaphors suggests that there is no need for such a possibility anyway: the process can be thought of as purely associative. At one point, Nietzsche claims explicitly that 'metonymy lies at the essence of synthetic judgment' (*PT* 152). Along the same lines, his insistence on the exclusive part played by the imagination in the formation of representations can be seen as an indirect attack on the notion of schematism. In the first *Critique*, the role of the imagination is limited to bridging the gap between sensibility and understanding by making sure that our categorial framework can apply to intuitions.¹³ Thus the purpose of the imaginative schemata (as determinate pure intuitions) is to enable the unification of sensory impressions under the rule of the pure concepts of the understanding. Here, Nietzsche lifts such limitations by removing the need for any kind of conceptualization.¹⁴ However there are also some strongly anti-empiricist elements in his account. For one thing, most empiricists, although they deny that experience is conceptually united, still grant that it has regularity. Thus for Hume our mental associations are not free: they are regulated, not by a priori concepts or forms, but by psychological principles such as resemblance, contiguity, or causality. Yet this is rejected by Nietzsche's insistence on the 'arbitrary' (*TL* 82) power of the imagination and the irregularity and lack of coherence of our perceptual metaphors. Secondly, both the notion of metaphor itself and the idea that the activity of the imagination is 'primal' suggest that there is nothing which would count as a raw sense datum, passively imprinted on the mind. On the contrary, for something to count as a perceptual content there is need for the interpretative activity of the imagination (which is why the relation between stimulus

¹² Amongst other things, one could argue that it is difficult to understand how 'images' could be formed without any conceptual input. Moreover, it would be difficult to distinguish between two different images if one could not avail oneself of such concepts as coexistence, succession, etc. I shall return to these issues later.

¹³ Cf. A136/B175: schematism provides the 'sensible conditions [the schemata] under which alone pure concepts of the understanding can be employed'.

¹⁴ One might be tempted to say that Nietzsche is shifting to an aesthetic perspective, according to which the imagination schematizes without concepts. However, even the third *Critique* requires the existence of a *sensus communis*, and the postulation that reflective judgements can be universal, two requirements which are clearly not endorsed by Nietzsche's account of perceptual metaphors here.

and image is not merely causal). The very disjunction of the spheres of nervous impulse and image suggests that a pure nervous stimulus could not register on the mind nor become significant for us unless it was interpreted by the imagination. As Nietzsche puts it, a stimulus must be 'seen as red, another as blue', or 'heard as a sound' (*TL* 87, my emphases). Per se, it is neither blue nor red: it is nothing to us. The 'artistic' power of the imagination thus lies in its ability to transpose nervous stimuli into images which are meaningful for us.¹⁵

So the first way of understanding Nietzsche's brief comments on the metaphorical nature of perception is to construe them as an independent description which does not rely on any ontological assumptions: it is purely based on an internal analysis of the genesis of our representations which denies the possibility of synthetic a priori judgments without, however, endorsing an empiricist view of experience. Yet Nietzsche immediately (and rather confusingly) shifts the discussion to another, more metaphysically loaded, level: his analysis of perception is now itself taken as a metaphor illustrating the inscrutable relation of human knowledge to what is seen as the essence of things. The focus moves imperceptibly from the idea that the genesis of perceptual experience does not warrant the possibility of objective empirical knowledge to the claim that it is impossible for our representations to correspond to the true nature of things anyway. The argument here is not that objectivity is a priori impossible: it is that, even if it was possible (and thus if we could have true knowledge of the empirical world), still our representations would fail to describe the world as it is. We would be in a situation analogous to that of a deaf person contemplating Chladni's figures: seeing the waves produced by sound on sand would get us no closer to understanding what sound is, and would just leave us puzzled.¹⁶ Thus 'a painter without hands who wished to express in song the picture before his mind would, by means of this substitution of spheres, still reveal more about the essence of things than does the empirical world' (*TL* 87). Nietzsche seems to lean towards a phenomenalist view of transcendental idealism,¹⁷ and to claim that there is a radical disconnection between the world of empirical representations and the noumenal world of essences. He suggests two reasons for this: first, he rejects the (Schopenhauerian) idea that there could be a causal relation between the thing in itself and the stimuli we receive: 'the further inference from the nerve stimulus to a cause outside of us is already the result of a false and unjustifiable

¹⁵ As we shall see, this anti-empiricist spirit remains present until the end of Nietzsche's work: thus, 'everything of which we become conscious is arranged, simplified, schematized, interpreted through and through' (*WP*, §477), or again 'everything that enters consciousness as a "unity" is already tremendously complex' (*WP*, §489). However, it should be noted that it sits awkwardly with Nietzsche's anti-conceptualism in his early account of perceptual metaphors. Further along in *TL*, and also in his later work, Nietzsche holds that judgement plays a bottom-down part in the formation of experience.

¹⁶ 'Perhaps such a [totally deaf] person will gaze with astonishment at Chladni's sound figures: perhaps he will discover their causes in the vibrations of the string and will now swear that he must know what men mean by "sound"' (*TL* 82).

¹⁷ Thus 'from the very beginning we see the visual images only *within ourselves*; we hear the sound only *within ourselves*. It is a big step from this to the postulation of an external world' (*PT*, §144).

application of the principle of sufficient reason' (TL 82). Between appearances and things in themselves, as between stimulus and image, there is only a metaphorical, subjectively established relation: 'in the same way as the sound appears as a sand figure, so the mysterious X of the thing in itself first appears as a nerve stimulus, then as an image, and finally as a sound' (TL 83). Secondly, even if there was a causal relation between thing in itself and appearances, still this would give us no understanding of the thing in itself: if we thought so, we would be as mistaken as a deaf person who, having 'discovered the cause [of the sand waves] in the vibrations of the string', was now to 'swear that he must know what men mean by "sound"' (TL 83). Nietzsche therefore concludes that 'nature is acquainted with no forms and no concepts, but only with an X which remains inaccessible and undefinable for us' (TL 83).

This second, better known, line of argument, which is based on the impossibility for human knowledge to apply to the noumenal world, is both close to and substantially different from Kant's own position. Clearly, Nietzsche is committed to the ontological thesis that there is a way in which things are in themselves, independently from us, and to the epistemological claim that such things, considered from a transcendental perspective (i.e. that of the bracketing of epistemic conditions), are by definition unknowable for us. However, the latter was not a problem for Kant's epistemology: the fact that we cannot know anything of the essence of things considered independently of epistemic conditions is a consequence of transcendental idealism, not an argument against the possibility of objectivity in the empirical world. Yet Nietzsche introduces another requirement, that of metaphysical correspondence: he thinks that, in order to count as knowledge, our statements must adequately reflect the way things are independently of us. As this is by definition impossible, Nietzsche concludes that human knowledge is *necessarily* false (the so-called error theory). Like Kant, he is what I shall call an ontological metaphysical realist, in that he thinks that it makes sense to talk of things having an essence independently from the way they appear to us. But unlike Kant, he is a disappointed *epistemological* metaphysical realist: he thinks that human knowledge *should* be able to capture the essence of things in themselves, and argues from its inability to do so to the radical impossibility of any sort of knowledge, be it empirical or noumenal. The way in which he argues against Kant in the following quotation is particularly indicative of his commitment to a (non-Kantian) conception of truth as metaphysical adequation: 'there is to be sure a vicious circle here: if the sciences are right, then we are not supported by Kant's foundations. If Kant is right, then the sciences are wrong' (PT, §84). What underlies the argument is the assumption that for the sciences to be right is to capture the essence of things considered in themselves. Thus if they are right, their being right is directly grounded by virtue of metaphysical correspondence. One can be a happy metaphysical realist, without any need for transcendental idealism or 'Kant's foundations'. And if Kant is right, i.e. if by definition things in themselves are beyond the scope of human knowledge, then the sciences are 'wrong' because they fail to capture the essence of reality. Of course, Kant would say that only if he is right can we be assured that the sciences are right too. The fact that

Nietzsche does not see this is indicative of his implicit commitment to metaphysical correspondence, which is the first ground for his error theory.

As should begin to emerge, this second line of argumentation does not seem particularly strong. For one thing, it rests on a misreading of Kant's intentions in the *Critique of Pure Reason*, and consequently on the sneaking-in of a premise (the metaphysical correspondence requirement) which a proper construal of the critical project would have shown as undesirable. Correlatively, it presupposes an undue generalization from the impossibility of empirical knowledge to represent the in-itself to the impossibility of empirical knowledge tout court. But it does not follow from the fact that human knowledge must fail the metaphysical correspondence requirement that it must also fail to adequately reflect *empirical* states of affairs. Moreover, even if one granted the applicability of the metaphysical correspondence requirement to all forms of knowledge, it still would not obviously follow from the transcendental idealist's claim that things or states of affairs are unknowable from a transcendental perspective that empirical knowledge could not correspond to things as they are in themselves, although of course there would be no way we could know that such a correspondence exists, let alone ground it. In other words, our knowledge could happen to be metaphysically true (in an adequationist sense), although we could not justify why it is true. As is well-known, this is a particularly thorny issue in Kant studies (the so-called neglected Trendelenburg alternative). Nietzsche himself is aware of this possibility, and actually criticizes Kant twice for having excluded it: thus 'against Kant, it must be further objected that even if we grant all his propositions, it still remains entirely *possible* that the world is as it appears to us to be' (*PT*, §84). Equally, 'we should not presume to claim that this contrast [between individual and species] does not correspond to the essence of things: that would of course be a dogmatic assertion and as such, would be just as indemonstrable as its opposite' (*TL* 83–4). Nietzsche thus rebukes Kant for not having been *critical enough*: he should have seen that the claim that empirical knowledge cannot correspond to metaphysical states of affairs is an unwarranted synthetic a priori judgement.¹⁸ Yet strangely enough, he does not seem to be aware that the criticism would apply equally well to his own falsification thesis, and that by the same token, he is unjustified in claiming that our knowledge necessarily fails to capture any features of things in themselves.¹⁹

¹⁸ There are ways of at least trying to make the claim good, in particular Allison's argument that it follows *analytically* from Nietzsche's considerations about space and time being conditions of possibility of representation that they cannot apply to the things in themselves, and that therefore the claim that things in themselves are not spatio-temporal is not synthetic but analytic. But as Nietzsche himself does not consider the issue, and seems to endorse the Trendelenburg alternative, I shall leave this issue aside.

¹⁹ In Nietzsche's defence, it could perhaps be said that, even if our statements happened to correspond to the in-itself, the very conditions under which the problem is set would prevent them from *counting* as knowledge. Interestingly, the situation evoked here seems to anticipate the Gettier problem: as it is well known, one of Gettier's examples is that of someone who, upon seeing a robot-dog on a lawn, says: 'there is a dog over there'. As the perceived dog is not a real dog, the statement is false (although justified), and thus cannot count as knowledge. Gettier then introduces the additional premise that there is a real dog behind the robot dog, hidden from the speaker by a bush. Gettier's conclusion is that, although the statement

At this point, it might be useful to summarize our findings, with a view to identifying the play of transcendental and naturalistic elements in the early Nietzsche's thought. I have suggested that he has two distinct arguments against the possibility of objectivity: the first one concerns the genesis of representations and is highly ambiguous in that it combines both anti-transcendentalist and anti-empiricist aspects. The second rests on his endorsement of a robust form of transcendental idealism and of the non-Kantian requirement of metaphysical correspondence. However, we are not at this point in possession of all the elements in the picture, and must therefore consider the additional arguments offered by Nietzsche in *TL* before drawing any definitive conclusion.

Temporarily leaving aside the issue of metaphysical correspondence, Nietzsche returns to his analysis of the genesis of representations, but this time from a diachronic perspective. The new and important element is that his former analysis of perception as metaphorical is now recontextualized within a genealogy which acknowledges that conceptualization is now needed, and tries to explain why. Nietzsche is concerned with showing the ways in which our imagination, while originally unconstrained, has progressively been forced to operate in a fixed manner. This pre-genealogical account is distinctively Humean in that it insists on the part played by social needs and habits in the process, and thus provides naturalistic reasons for the taming (and the forgetting) of our 'primal' metaphorical power. Starting from the implicit premise that it is necessary for human beings to live in society (presumably for Rousseauist reasons having to do with the scarcity of natural resources and the natural vulnerability of the species), Nietzsche's account is developed in the following steps. (a) For social order to be possible, there is a need for its members to be able to be truthful. It is a 'duty which society imposes in order to exist'²⁰ (*TL* 84): otherwise it would be threatened with chaos and could not maintain itself. (b) '[T]o be truthful means to employ the usual metaphors' (*TL* 84), i.e. (for Nietzsche) concepts, presumably so that statements are communicable,²¹ verifiable, and justifiable. If there was no common reference, or no stability in the relation between meaning and reference, then no societal life would be possible. (c) Yet such concepts are themselves deemed false, by virtue of the process of abstraction whereby they are obtained. (d) Consequently, we have the 'duty to

then would turn out to be true, it would still not count as knowledge. In the same way, it could perhaps be argued (although Nietzsche doesn't) that if our statements, which are deemed by Nietzsche empirically false because of the lack of metaphysical correspondence and thus are not considered as items of knowledge, turned out to be metaphysically true, but for reasons which are unavailable to us, they still could not count as knowledge. It would just be a case of epistemic luck.

²⁰ The reference to 'duty' in this context could be seen as ironically reminiscent of Kant's arguments both in the *Groundwork* in 'Of the So-called Right to Lie'. Polemically, Nietzsche sees the empirical origin of what Kant understands as an a priori moral duty in the need for social order and cohesion.

²¹ Cf. *WP*, §569: 'communication is necessary, and for there to be communication something has to be firm, simplified, capable of precision (above all in the so called *identical* case). For it to be communicable, however, it must be experienced as adapted, as "recognisable". The material of the senses adapted by the understanding, reduced to rough outlines, made similar.'

lie according to a fixed convention, to lie with the herd and in a manner binding on everyone' (TL 84).

Clearly, a lot hinges on step (c). What is the ground for saying that conceptualization is necessarily a form of falsification? And what exactly is being falsified? Two main reasons can be inferred from Nietzsche's text. The first one is that conceptual knowledge fails to fulfil the metaphysical correspondence requirement. This is probably Nietzsche's weakest argument in that it is the least specific (it would apply equally to the primal stream of metaphors): from this perspective, it is difficult to see why fixed metaphors would be, as Nietzsche clearly thinks they are, *worse* than unfixed ones. The second reason is reminiscent of the empiricists' criticism of the formation of abstract ideas: it is the claim that concepts are formed a posteriori, through an impoverishment of the original perceptual metaphors. The process can be detailed in two steps: first, the original, singular images are 'universalized . . . into less colourful, cooler concepts, so that [man] can entrust the guidance of his life and conduct upon them' (TL 84). Such universalization is in turn made possible by the forced equalization of individual differences. Thus whereas 'each perceptual metaphor is individual and without equals and is therefore able to elude all classification' (TL 84–5), 'every concept arises from the equalisation of unequal things: just as it is certain that one leaf is never totally the same as another, so it is certain that the concept "leaf" is formed by arbitrarily discarding these individual differences' (TL 83).²² Secondly, we forget about this genesis and take concepts to refer to real entities in the world. Thus the formation of universals, conjoined with the arbitrary singling out (the 'arbitrary assignment', TL 82) of general properties such as 'green' which are implicitly referred to substances as their bearers, gives rise to the illusion of a world of mind-independent empirical entities: thus we 'proceed from the error of believing that we have these things immediately before us as mere objects. We forget that the original perceptual metaphors are metaphors and take them to be the things themselves' (TL 86). There is therefore a double falsification at play: first, the original individual perceptual metaphors are solidified into concepts; secondly, this process is forgotten and we become naive realists, believing in the independent existence of entities picked out by these concepts. Although Nietzsche does not say it explicitly, presumably this allows for the attribution of a truth value to our statements, since such entities can then serve as referents and thus make it possible to verify our statements.

This is the first appearance of a theme (namely that concepts, and therefore conceptual knowledge, necessarily falsify) which offers a second ground for Nietzsche's error theory. Note that, like his first views about the metaphorical character of perception, it is independent from his commitment to transcendental idealism, and therefore it will

²² Nietzsche presupposes that our ability to distinguish different objects is non-conceptual: yet one could argue that in order to identify a leaf as such in the first place (and thus to be able to tell two leaves apart) one already needs the concept of a leaf, and therefore that it is presupposed by the very process of which it is described as result.

not be affected by his ulterior rejection of the thing in itself. Because of its importance, it is worth examining this early version of the conceptual falsification thesis in a little more detail. Perhaps the first thing to say is that, for all its apparent simplicity, it is deeply ambiguous in that the nature of the falsified varies implicitly: in the quotations already given, Nietzsche first identifies it with 'each original perceptual metaphor', but then goes on to say that 'every concept arises from the equalisation of unequal *things*' (my emphasis). The example of the leaf confirms that he is now talking about the equalization of the differences between various *entities*, not perceptual metaphors (thus he talks about it being certain that 'one leaf [not a *perceived* leaf] is never totally the same as another'). In the same way, his criticism of our picking of properties as arbitrary presupposes a similar realist view of the falsified: thus 'we speak of a "snake": this designation touches only upon its ability to twist itself and could therefore also fit a worm' (TL 82). This would suggest that Nietzsche himself occasionally falls prey to the illusion he criticizes, namely that of the belief in the existence of mind-independent entities which pre-exist our perception of them.²³

Given the general focus of his argument in *TL* on perceptual metaphors, let's assume that what Nietzsche really means here is that concepts falsify our original imaginative metaphors. In this case, the falsification thesis assumes not only that we have from the start the ability to form many perceptual metaphors, but that such metaphors can be said to differ from one another. Otherwise, the idea of original '*singular*' images would make no sense, and nor would that of a subsequent conceptualization as equalization of differences in perceptual content—all we would have is a stream of undifferentiated perceptions which thus could not be individuated enough to be falsifiable. But Nietzsche is clear that each perceptual metaphor is 'a unique and entirely original experience' (TL 83). Then the question arises of whether we are originally aware *that* our perceptual metaphors are different from each other (as opposed to just having different perceptual metaphors). Nietzsche is not clear on this matter, but it seems that there are two possibilities. First, if the equalizing process of conceptualization is seen as conscious and intentional, then it does require that each perceptual metaphor should be perceived as different from the others from the start. Otherwise it simply would not be possible to compare them in order to form the smallest possible common denominator (the concept). In this case, the existence of such awareness reinforces the anti-empiricist aspect of Nietzsche's account of perception in the sense that the perception of difference as such cannot be accounted for by merely having different perceptual metaphors.²⁴ Yet by the same token, one could argue (along Kantian lines) that the idea that it is possible to differentiate thus between perceptual particulars without any conceptual input at all is dubious. It would seem that in order to have a sense of

²³ I shall return to this point, and suggest an interpretation of the development of the error theory as a warning against precisely this illusion, rather than against the possibility of knowledge in general.

²⁴ As it is well known, Kant makes a similar objection to Hume about succession: having a perception of succession, as we clearly have, is not reducible to having successive perceptions.

difference, one needs a minimal conceptual framework, including such general concepts as, for example, quantity, unity, or succession. Otherwise perceptual metaphors would merge into one another and be undistinguishable. This picks up on a problem I flagged earlier, namely the idea that it is difficult to understand how the imagination could, without the use of any concepts, have the formative and individuating activity that Nietzsche attributes to it. Although their content may not be fully articulate, the original perceptual metaphors could not stand as distinct 'images' without the use at least of categorial concepts.

One way out of this problem is to move to the second possibility alluded to and to construe the equalizing process that results in the formation of concepts as *unconscious*. In this case there would be no requirement for us to be aware of the differences between perceptual metaphors. Motivated by social need, we could simply react insensibly to such differences in a manner which progressively reduces them so that we end up with conceptually structured and thus communicable representations. This would be a fully naturalistic account of the formation of concepts, in line with Nietzsche's insistence on the role of habit and need in their genesis. However it sits uneasily with his non-empiricist account of the part played by the imagination in the formation of perceptual metaphors, and also with other claims which suggest that the formation of the conceptual edifice was intentional, deliberate, and conscious.²⁵ Furthermore, if there was no awareness of the difference in perceptions from the part of the perceiver, then it is difficult to determine the perspective from which they could legitimately be said to have been different. The claim can only be made from the third person, from the point of view of the philosopher (Nietzsche) who retrospectively identifies the differences in perception and diagnoses the equalization. But how can he be assured that such identification is legitimate, and that there really were differences in the first place? Not only is he describing a long-gone, mythical past, which he says has been 'forgotten' (TL 84, 86); he is also assuming that it is possible to describe a first-person experience (having the different perceptions, although without being aware of their differences) from a third-person perspective, which is made problematic by the private character of such perceptions.²⁶ It could be argued that he is relying on an analogy between his current perceptions and the ones attributed to the original, pre-conceptual humans. But *he* is clearly highly aware of perceptual differences. Moreover, the simple act of describing the process would transform the original experience, if there was one. So both the language of the description and the way in which Nietzsche experiences the world are now incompatible with the primitive experience he is trying to describe. In such a case, it becomes difficult to provide any argument for the claim that there was

²⁵ 'Man now places his behaviour under the control of abstractions. He will no longer tolerate being carried away by sudden . . . intuitions. First he universalizes all these impressions into less colourful, cooler concepts, so that he can entrust the guidance of his life and conduct to them' (TL 84, my emphasis).

²⁶ Remember that, in this hypothesis, there is no conceptual element in perception and no common language to provide such perceptions a public character.

a difference in perceptual content which would have been unconsciously equalized, because it is impossible to establish the truth of the premise. At best the whole reasoning can be seen as hypothetical, which is not the modality of Nietzsche's speech (it is assertoric).

So Nietzsche's original two lines of argumentation, namely his characterization of experience as 'perceptual metaphor' and his endorsement of transcendental idealism joined with the non-Kantian requirement of metaphysical correspondence, are now complemented by a proto-genealogical account of the ways in which conceptual use has become necessary. This seems to culminate in a rather Humean vision of experience: 'even the relationship of a nerve stimulus to the generated image is not a necessary one. But when the same image has been generated millions of times and has been handed down for many generations and finally appears on the same occasion every time for all mankind, then it acquires at last the same meaning for men it would have if it was the sole necessary image and if the relationship of the original nerve stimulus to the generated image were strictly a causal one' (TL 87). Having forgotten both the part originally played by the imagination in the formation of perceptual metaphors and the reductive process whereby concepts were acquired, we have 'convinced ourselves of the eternal consistency, omnipresence and infallibility of the laws of nature' (TL 87). *Prima facie*, both the genealogy and its conclusion seem to reinforce the naturalistic elements in Nietzsche's thought. However, they unexpectedly result in a substantially different and much more transcendently inclined account of the way experience is structured *now*, i.e. at the end of the historical process described in TL.

The main difference is this: whereas before Nietzsche denied the need of any kind of a priori framework constraining the formation of perceptual metaphors, he now states that 'the artistic process of metaphor formation with which every sensation begins in us *already presupposes* [these] forms and thus *occurs within them*' (TL 87, my emphasis). Such forms are identified as 'time and space, and therefore relationships of succession and number'²⁷ (TL 87). Other passages specify that 'space, time and the feeling of causality appear to have been *given along with the first sensation*' (PT, §80, my italics), or again that 'the perceived manifold *already presupposes* space and time, succession and coexistence' (PT, §140, my italics). Two things are worth noting: first, such claims set Nietzsche even further apart from an empiricist account of the formation of experience in that they implicitly reject the idea of a tabula rasa. While experience does rely on some form of sensory input, it also requires the existence of a priori elements. Each time we experience something, the forms are 'already' there, 'given along with the first sensation'. The forms do not cause the sensations (as we have seen, for Nietzsche it cannot even be said that nervous stimuli do). Correlatively, their anteriority is not chronological (it would make no sense to say that the forms exist before experience as such relations of anteriority or posteriority can only be defined from the perspective

²⁷ An implicit reference to the Axioms of Intuition, the Anticipations of Perception, and the Analogies?

of experience, i.e. from within the framework provided by such forms). Although they can be defined in isolation, they do not pre-exist experience, or exist independently from it. Nietzsche seems to think that they *shape* such experience (as opposed to experience merely happening in conformity with them): thus, because of them we 'bring to things' their properties (TL 87). This may implicitly motivate (or at least reinforce) his claim that the world as we perceive it cannot be the world of things in themselves, and thus that human knowledge by definition fails the metaphysical correspondence requirement. Secondly, Nietzsche does not seem to distinguish strongly, as Kant does, between the a priori forms of sensibility and the pure concepts of the understanding: time, space, causality, coexistence, and succession (neither of the last two being, strictly speaking, categories) are placed on an equal footing. This may be partially due to Schopenhauer's influence (since the latter included causality, along with time and space, in his definition of the principle of sufficient reason). It may also indicate that, in fact, Nietzsche does not think that it is possible to consider sensory content independently from some form of conceptualization (we have seen that this was a problem for his view of perceptual metaphors as relying purely on the imagination). Yet such conceptualization should not be seen as requiring a full-blown, verbal, and consciously articulate use of concepts. By placing causality, existence, and succession on the same level as time and space, Nietzsche seems to hold that perception requires minimal ordering, which cannot be provided by time and space alone. We have to be able to judge that some impressions happen after others in time, or coexist with them in space. As we shall see, this anti-empiricist aspect is reinforced in the later work.

Yet although they look transcendental in the Kantian sense, the conditions identified by Nietzsche are not truly so. The point of his genealogical account is to show how they have been *acquired*, and how the way we perceive the world consequently changed. Thus by opposition to the time when our imagination ran free, we are now '*forced to comprehend all things only under these forms*' (TL 87, my emphasis). Moreover, Nietzsche's proto-genealogy indicates that since conceptualization arose from need and was enforced by habits and forgetfulness (both of the singularity of the original perceptual metaphors and of the part played by the imagination), there is no *rational* validation for the necessity of these forms: 'we produce these representations [of time and space] in and from ourselves with the *same necessity* [i.e. presumably, natural necessity] *with which the spider spins*' (TL 87). The 'transcendental' elements in perception have an empirical genesis: they have their causal conditions of possibility in the rise of new practices (such as societal life), and their necessity is only relative to these practices. Thus although Nietzsche's account of the current way we experience the world is certainly not empiricist, it is not truly transcendental either. In order to bring out its specificity, it may be useful to refer to Mark Sacks's distinction between transcendental constraints and transcendental features (Sacks 2000, 211–18). The first indicates a 'dependence of empirical possibilities on a non-empirical structure' (Sacks 2000, 213). In such a case, the conditions of the constitution of experience are definable in isolation and in anticipation of what they determine (in the way the transcendental

organization of the faculties can be spelled out completely independently of experience in Kant, and in such a way that experience must conform to them). This makes it possible, at least in theory, to secure the possibility of objective knowledge completely, i.e. in a universal way. It would not make any sense to think of a change in transcendental constraints which would result in a change in experience; on the contrary, they are what allows us to think the possibility of succession (and thus of change) and to identify it in the empirical world.

By contrast, transcendental features 'indicate the limitations implicitly determined by a range of available practices . . . to which further alternatives cannot be made intelligible to those engaged in them' (Sacks 2000, 213). To claim that such features exist is to reject the empiricist (or naturalist) view that experience is the result of purely associative processes which ultimately depend on physical causes (such as changes in the environment, or in the neurophysiology of the brain). This accounts for their 'transcendental' aspect: they operate like transcendental constraints in that they delimit the structure that experience has to conform to in order to count as experience. Yet they can only be considered a priori from an artificially induced synchronic perspective, if one looks at experience at an instant *t* and inquires, in a decontextualized way, about the necessary structure of such experience. Ultimately they must be placed within a diachronic account (in this case, Nietzsche's proto-genealogical project) which shows that their genesis is dependent on changes in empirical practices (such as the need for truthfulness in *TL*). Thus, contrary to constraints, transcendental features can only guarantee the possibility of objectivity in a limited way, within the context of a certain range of practices. If the practices change, so will the transcendental features which ultimately are 'no more than a shadow of necessity cast by whatever practices are current' (Sacks 2000, 213). This, it seems, pretty much captures the spirit of Nietzsche's view: there are some apparently non-empirical conditions for knowledge, but these are ultimately dependent on empirical circumstances. On the one hand, philosophy must 'bring to light the preconditions upon which the process of reason depends' (*WP*, §487), such as the 'categories', 'Euclidean space' (*WP*, §515), or the 'a priori law of causality' (*KSA* 25 = *W I* 1. Spring 1885, 26 [74]).²⁸ Such preconditions work in a transcendental way insofar as their validity cannot be verified by empirical means (since such proof would presuppose their use), and because it is impossible to identify them with any particular set of practices (thus although *TL* suggests that they appeared along with new practices, they are not *identical* to these practices). Yet on the other hand, 'the mode under which we know and form knowledge is already itself part of our conditions of existence: but . . . this *factual* condition of existence may be only *fortuitous* and not necessary at all' (*KSA*: 25 = *W I* 1. Spring 1885, 26 [127]; the first emphasis is mine, the others Nietzsche's). Correlatively, such conditions are not rationally justified but grounded in belief: their necessity is psychological, and the truths that they provide are only 'conditional' (*WP*, §515).

²⁸ All the translations from the *Nachlass* are mine.

So while both Nietzsche and Kant are transcendently inclined in that they have a non-empiricist view of experience and agree that it necessarily presupposes the use of a spatio-temporal and conceptual framework, Kant sees the latter as a constraint, and Nietzsche as a feature. This has two consequences: first, it makes it even more difficult to tell whether Nietzsche should be seen as a naturalist or a transcendental philosopher. There is an intrinsic instability (or reversibility) to the notion of transcendental features: their existence can equally be seen as an argument for the transcendental philosopher (in that the constitution of experience conforms to a priori elements which are not accountable for through empirical psychology or neurophysiology) *and* for the naturalist (in that such elements are not a priori in the strong sense of constraints, as they ultimately depend on contingent practices which could themselves be accounted for in a naturalistic way, especially if one is a 'soft' naturalist). One can insist either on the anti-empiricist thrust of such an account of experience, or on the fact that the a priori, being empirically rooted, is neither universal nor necessary in the non-causal, non-psychological manner required by transcendental constraints. Furthermore, Nietzsche's proto-genealogy suggests both that the possibility of knowledge is contingent on the existence of a specific framework, and that such a framework is revisable if the conditions of experience change (and, in fact, has already been revised), something a true transcendental philosopher would not accept. Secondly (and correlatively), a similar ambiguity carries over to the relation between transcendental features and objectivity.

The advantage of a transcendental approach over a purely empiricist or naturalist one is that, if it succeeds, can guarantee a priori the possibility of an agreement between experience and its objects. On such an account, experience can stop (or not happen) but not fail to conform to its a priori conditions. The possibility of such conformity is preserved by the notion of transcendental features, but in a much more limited way: they guarantee that at a given time, and under certain empirical conditions, there will be harmony between the conditions under which we know and the structure of the objects we can know. But should the empirical context change, other forms of knowledge and other objects will appear (thus in *TL*, causally related spatio-temporal objects in opposition to the original perceptual metaphors). Objective knowledge is now possible, but it isn't truly universal. Whereas the singular and randomly formed character of perceptual metaphors made it impossible for our representations to be regulated by any rules and for us to have a shared perceptual world, now the use of both a priori forms and concepts can in principle guarantee the possibility of agreement between the conditions under which we perceive and the perceived world. Nietzsche is well aware of this: 'all that conformity to law, which impresses us so much in the movement of the stars and in chemical processes, coincides at bottom with those properties which we bring to things' (*TL* 87). Or again, 'in all things we comprehend nothing but these forms. For they must all bear within themselves the laws of number' (*TL* 87). Yet because of his commitment to transcendental idealism and to the metaphysical correspondence requirement on the one hand, and of the

falsification thesis on the other, this tentative opening for objectivity aborts immediately: thus 'we presuppose that nature behaves in accordance with such a concept. But in this case first nature and then the concepts are anthropomorphic . . . The essence of things does not correspond to this: it is a process of knowledge which does not touch upon the essence of things' (*PT* 150). But should his commitment to ontological metaphysical realism and his error theory disappear, this would in principle leave some scope for a relativized form of objectivity within the framework of naturalized transcendental features.

III

I now want to look at the ways in which the themes set in place by *TL* develop in Nietzsche's middle and late periods. As previously, what interests me most is the interaction between naturalistic elements, transcendental aspects, and their ontological assumptions. I shall therefore focus on three main issues: (a) Nietzsche's alleged rejection of transcendental idealism and the sort of ontological commitments (if there are any) that follow from it, (b) his reinterpretation of transcendental features as perspectival 'conditions of life', and (c) whether it can be rescued from the threat posed by the error theory.

As pointed out by Clark and Leiter, from *HH* onwards Nietzsche seems increasingly distrustful of ontological claims, and presents two distinct arguments. First, in a quasi-Allisonian way he identifies the possibility of the existence of a noumenal realm as the analytic correlate of the notion of perspectival conditions. Thus, 'it is true, there could be a metaphysical world: *the absolute possibility of it is hardly to be disputed*. We behold all things through the human head and cannot cut off this head; while the question nonetheless remains of what of the world would still be there, if one had cut it off [i.e. if epistemic conditions were bracketed]' (*HH* i, §9, my emphasis). Yet Nietzsche argues that, although such existence can be legitimately conceived, it is of no importance whatsoever to us because by definition we cannot know anything of such a world which would make a difference to our lives: 'one could assert nothing at all of the metaphysical world except that it was a being-other, inaccessible, incomprehensible being-other; it would be a thing with negative qualities. . . . Knowledge of it would be . . . more useless even than the knowledge of the chemical composition of water must be to the sailor in danger of shipwreck' (*HH* i, §9). Thus Nietzsche seems to implicitly move from a robust interpretation of transcendental idealism to a more deflationary one, focused on the noumenon as a purely negative concept, and then to make the further, pragmatic point that we shouldn't concern ourselves with such a possibility because it is irrelevant to us. The underlying idea seems to be that, while the robust interpretation of transcendental idealism did matter because of its nefarious moral implications (and thus had to be rejected), this deflationary view can stand because it is so empty as to be harmless. To this argument, Nietzsche adds a second, more ambiguous one: 'as though knowledge

here got hold of its object purely and nakedly as the “thing in itself”, without any falsification on the part of either the subject and the object! But . . . the “thing in itself” involve[s] a *contradiction in adjecto*’ (BGE i, §16, Nietzsche’s emphasis). The first part of the quotation suggests that the idea that we can know the thing in itself is contradictory (which certainly follows from its concept, unless such knowledge is inferred analytically from the definition of such a thing); but the conclusion suggests that it is the very *notion* of a thing in itself which is contradictory. Clark develops the contradiction as follows: ‘we can have no conception, or only a contradictory one, of something that would be independent of all knowers, and therefore of all conceptualization, because to conceive of something is to conceive of it as satisfying some description or other, which is to think of it as being conceptualizable in some way or other’ (Clark 1990, 46–7). Yet this gloss only works at the cost of introducing implicitly a *different* conception of the thing in itself than the one indicated in *HH*. There Nietzsche suggested that the metaphysical world is merely the correlate of the bracketing of the transcendental features that are relevant to *us* (what remains if one ‘cuts our head off’), not of the bracketing of *any* epistemic conditions. To recast this in his later vocabulary of perspective, the first idea is that things in themselves are things considered independently of our perspective; yet a contradiction in adjecto can only be made good if things in themselves are thought of as entirely extra-perspectival. Only then does it become possible to argue, as Clark does, that the very existence of the definition presupposes that of a perspective from which it originates, and therefore that there is a latent contradiction between the definiens and the nature of the definiendum.

Prima facie, later Nietzsche makes a move of that sort when he claims that a ‘thing’ (both in its traditional sense, as substance, and in its more Kantian understanding as thing in itself) is nothing more than ‘the sum of its effects’ (*WP*, §551). According to this logic, ‘the “thing in itself” is indeed nonsensical. If I remove all the relationships, all the properties, all the activities of a thing, the thing does not remain over’ (*WP*, §558). As pointed out by Poellner, it is dubious whether this argument would truly apply to Kant;²⁹ at any rate, it is clearly designed to deny the possibility of strong metaphysical ontological realism by invalidating the claim that there is a way in which things are *in themselves*. Thus ‘the “in-itself” is even an absurd conception: a “constitution in-itself” is nonsense; we possess the concept “being”, “thing”, only as a relational concept’ (*WP*, §583). However, it does not follow from this that it would make no sense to ask again, as Nietzsche himself did in *HH*, whether some *X* would remain if *our* perspectival conditions were bracketed. Although the idea that such *X* could have a self-standing essence, definable independently from any perspective, has just been rejected, and although *we* couldn’t say anything positive about it, there could be *other* perspectives from which such an *X* could be considered. Most interestingly, it is this

²⁹ Poellner 2001, 96 n. 20.

very possibility that a passage from *The Gay Science* considers, in terms which are strikingly similar to those of *HH*:

how far the perspective character of existence extends or indeed whether existence has any other character than this . . . —that cannot be decided even by the most industrious and most scrupulously conscientious analysis and self examination of the intellect: for . . . we cannot look around our own corner. . . . But I should think that today we are at least far from the ridiculous immodesty that would be involved in decreeing from our corner that perspectives are permitted only from this corner. (*GS* v, §374)

Thus the critical project (the ‘scrupulously conscientious . . . self examination of the intellect’) can only carry us to the limits of our own perspective: we cannot transcend the conditions by which our experience is structured. However, such a critique also reveals that our perspective is not the only possible one, although it is the only one which we can understand. Thus ‘let’s say that [what “we believe in most, all the *a priori*”] presents itself as a condition of existence for our species—a sort of fundamental hypothesis. This is why other beings could make other hypotheses, four dimensions for example’ (*KSA* 25 = *W* I 1. Spring 1885. 25 [307]). It is conceivable that there could be a plurality of perspectives, each with its own kind of *a priori* conditions (thus ‘some beings might be able to experience time backwards, or alternatively forward and backward,’ *GS* v, §374), which are incommensurable: ‘it is likely that there are numerous kinds of intelligences, but each has its *own laws* which make it *impossible* for it to *represent* another law’ (*KSA* M III 1. Spring–autumn 1881 11[291]). The existence of these other perspectives is enough to warrant a minimal ontological commitment to the possibility of there being some *X* considered independently from us, but not from any perspective. Most importantly, however, Nietzsche does not say that there *is* such an *X*, only that perspectivism allows us to conceive of such a possibility. Thus his position in his later work is more subtle than is usually allowed for: although he rejects the notion of things in themselves as overdetermined, he does not make the (equally strong) claim that, if the conditions that structure our experience were bracketed, *nothing* would remain. By contrast, his criticism of the thing in itself and the fact that he acknowledges the possibility of a plurality of perspectival conditions (as opposed to single, universal constraints) lead him to remain agnostic about ontological commitments. Importantly, this removes one of the grounds of his error theory: if whatever remains if our epistemic conditions are bracketed has no ‘constitution in-itself’, then metaphysical correspondence cannot be a requirement.³⁰

This subtle reflection on the ontological commitments of transcendental idealism is accompanied by a deepening of Nietzsche’s understanding of transcendental features, now reinterpreted as part of the perspectival conditions that make possible human

³⁰ Given the incommensurability of perspectives, it would not make sense either to require that the empirical knowledge we form should be valid cross-perspectivally. The implicit correlate of Nietzsche’s multi-perspectivism seems to be a form of empirical pluri-realism whereby different entities will be picked out according to different sets of transcendental features. Thus instead of ‘the [metaphysical] true world’ we should ‘assert the existence of “x” worlds . . . But this has never been asserted’ (*WP*, §586).

experience and life.³¹ Thus 'the belief in the truth [of synthetic a priori judgements] is necessary, as a foreground belief and visual evidence belonging to the *perspective optics of life*' (BGE i, §11, my emphasis). The existence of transcendental features is the reason why although each empirical individual's particular perspective may vary depending on this individual's idiosyncrasies, interests, and situation, all the objects that are singled out by human perspectives will share a few common, structural traits—in particular, being spatio-temporal, having an objective order of succession, being interconnected by the law of causality.³² Yet just as in his early work, Nietzsche seems uneasy with this idea. This dissatisfaction is expressed by the many statements that suggest that our knowledge, although grounded in unavoidable beliefs, might nevertheless be or even is an 'error'.³³ Such a claim is as ambiguous as it is problematic. If read literally, it threatens Nietzsche's own position in three different ways: first, as seen by most commentators, it creates a potentially fatal self-reference problem. Secondly, it seems to render Nietzsche's reinterpretation of Kant and his naturalization of transcendental constraints into features rather pointless as it makes the resulting epistemology incapable of accounting even for the possibility of limited objective representations. Thirdly (as we shall see), a literal interpretation of the error theory would require us to reintroduce precisely the sort of substantive ontological commitments from which Nietzsche tried to distance himself. Given these difficulties, the temptation is strong either to dismiss it, as Clark and Leiter suggest, or to minimize it as Nehamas does (Nehamas 1985, 51).³⁴ I shall, however, resist both temptations and suggest that the error theory should not be taken literally but seen as a deliberately hyperbolic warning against naive forms of realism and of transcendentalism, which in turn serves to reinforce the importance of Nietzsche's own naturalization of the transcendental.

Before developing this argument, let us take a closer look at the literal interpretation of the error theory: it is fuelled by two separate arguments both of which deepen Nietzsche's naturalization of the transcendental. As seen by Stack, the first one rests on his use of evolutionary psychology. While he accepts that our experience presupposes certain a priori conditions without which it couldn't be the experience it is, Nietzsche

³¹ Although no human perspective would be possible without the use of transcendental features, there is more to the notion of perspective than just such use. As indicated by Poellner (following Leiter), knowing an object from a perspective also means knowing it 'from the standpoint of particular interests and needs' (Poellner 2001, 99). Thus although each specific perspective will by definition presuppose the use of the transcendental features prevalent at that particular time, the sort of objects it focuses on will depend on the interests of the representing individual.

³² Cf. Poellner 2001, 88–98. All spatio-temporal objects have characteristics that mark them as represented (what he calls the 'essential representation-dependence thesis').

³³ For the weaker version of the error theory, see e.g. GS iii, §121 ('Life is no argument. The conditions of life might include error'); WP, §483 ('a belief can be a condition of life and nonetheless be false'), §487 ('a belief, however necessary for the preservation of a species, has nothing to do with truth'), §497 (cf. main text), §515 (cf. main text). For a stronger version, see HH i, §11, BGE §4, §24; TI, 'Reason in Philosophy', §5, WP, §493 ('Truth is the kind of error without which a certain species could not live'), §512 ('the will to logical truth can be carried through only after a fundamental falsification of all events'), §517.

³⁴ Thus the error theory would just express the lack of a God's eye perspective.

now gives a Darwinian³⁵ twist to such conditions: they are ‘conditions of life,’ in the dual sense that they have evolved with our form of life, and are required for such life to maintain itself and develop. Thus ‘if we establish what is *necessary* according to our current way of thinking, we have proved . . . only what “makes possible” our existence by virtue of experience—and this process is so ingrained that to try and turn thought away from it is impossible. Any *a priori* is located there’ (KSA M III 1. Spring–autumn 1881 11[286], Nietzsche’s emphasis). So while Kant started from experience and asked about the non-empirical conditions that ‘made it possible’ in the transcendental sense, Nietzsche recontextualizes such conditions within an evolutionary perspective and suggests ironically that having a certain form of experience (which does presuppose these ‘*a priori*’ conditions) is what ‘makes possible’ (in the causal sense, this time) human life, and was therefore evolved (along with its *a priori* conditions) so as to preserve and enhance such life. To put it plainly, we have become ‘hard-wired’ to use ‘*a priori*’ concepts and forms and to believe in the reality of the resulting objects; but the existence of such hard wiring is no guarantee for the truth of the resulting judgements. Thus ‘it is high time to replace the Kantian question, “how are synthetic *a priori* judgments possible” by another question, “why is belief in such judgments necessary”—and to comprehend that such judgments must be *believed* to be true, for the sake of the preservation of creatures like ourselves; though they might, of course, be *false* judgments for all that!’ (BGE i, §11, Nietzsche’s emphasis). Another passage develops a similar idea: ‘the law of *a priori* causality—that it is *believed* may be a condition of existence for our species; for all this, it is not *proved*’ (KSA 25 = W I 1. Spring 1885. 26 [74], Nietzsche’s emphasis).³⁶ The *a priori* conditions of possibility of experience are ultimately ‘conditions of existence’: their necessity is only relative to our need to believe in them, and their ‘truth’ only reflects the empirical conditions they result from. However, this argument is not decisive: it does not follow from this Darwinian naturalization of the transcendental that our knowledge is *false*—and in many passages, including all the quotes given here, Nietzsche himself is careful to only suggest that it *might* be the case. All that follows is that it is not true in the absolute, universal sense that would be guaranteed by the existence of transcendental constraints.

The second argument for the literal interpretation of the error theory is grounded in a further naturalization of the transcendental, this time through an analysis of language which naturalizes the Kantian categories. Nietzsche does not dispute that such primary concepts as ‘substance’ or ‘causality’ are presupposed by the way we experience the world. Yet contrary to Kant (and in line with the spirit of his early argument in *TL*), he does not view such concepts as truly *a priori*: he traces their genesis to the existence of specific grammatical features present in all Indo-European languages, in particular

³⁵ As argued by Richardson (2004), most of the time Nietzsche is *de facto* in agreement with Darwin, and only rejects vulgarized versions of Darwinism.

³⁶ Cf. also WP, §497: ‘the most strongly believed *a priori* “truths” are for me—*provisional assumptions*, e.g. the law of causality . . ., so much a part of us that not to believe in it would destroy the race. But they are for that reason truths? What a conclusion! As if the preservation of man was a proof of truth.’

the subject/object structure and the active/passive modes (from which are respectively derived the notions of subject and substance on the one hand, and of causality on the other). Thus 'positing as "true *a priori*" our belief in the concept of substance . . . is simply a formulation of our grammatical custom which adds a doer to every deed' (WP, §484); in other words, 'The concept of substance is a consequence of the notion of the [grammatical] subject' (WP, §485).³⁷ More generally, we are governed by our 'faith in grammar' (TI vi. 5; cf. also TI iii. §5). Nietzsche's point is not merely that whenever we *speak* of 'things' or 'subjects' or 'causes' we project onto reality grammatical structures which may be alien to it. The idea is that the very way we *experience* the world and individuate objects (prior to any verbal articulation), being structured by the categories, is ultimately shaped by the grammar of a specific and contingent set of languages. 'Our thinking itself involves this belief (with its distinction of substance, accident; deed, doer, etc.). To let it go means: being no longer able to think' (WP, §487). In a way that anticipates the linguistic turn in analytic philosophy, Nietzsche argues that the structure of our language shapes our perception of the world. We perceive what we take to be independently existing 'things', interacting causally with each other; but such perception only reflects the fact that our experience is made possible by concepts which themselves result from the unconscious reification of grammatical structures. Thus 'we have arranged for ourselves a world in which we can live—by positing bodies . . . , causes and effects, motion and rest, form and content: without these articles of faith nobody now could endure life' (GS iii, §121). Two points are worth noting: first, this naturalization of the categories does not yield per se the conclusion that our knowledge is erroneous, but merely that its scope and nature are dependent on the kind of linguistic conditions that determine it. The fact that we only perceive causally interacting things does not mean, per se, that our perception is *wrong*: it just means that it is dependent on our use of causality as a category, which itself is due to a specific characteristic of our grammar. Secondly, in many quotes what Nietzsche seems to object most to is not so much our use of grammar (and thus the existence of transcendental features), which is unavoidable, as the fact that we are unaware of the assumptions and consequences of such use. As we shall see, this may point towards a more fruitful way of interpreting the error theory.

Still, there is no doubt that Nietzsche's naturalization of the transcendental leads him to express skepticism about human knowledge. Yet as has been indicated, none of the arguments provided yield per se the conclusion that all human knowledge is erroneous. For the literal interpretation of the error theory to validate such a conclusion, two additional premises are needed, namely (a) that the world is *otherwise* than the way we construe it, and (b) that our knowledge fails to correspond to it. Whether this can be made good depends, in turn, on how premise (a) is construed, i.e. on the

³⁷ Cf. also WP, §631: 'The separation of the "deed" from the "doer", . . . of the process from something which is not process but enduring substance, thing, body, soul, etc.—the attempt to comprehend an event as a sort of shifting and place—changing on the part of a "being", of something constant: this ancient mythology established the belief in "cause and effect" after it had found a firm form in the functions of language and grammar'.

nature of the allegedly falsified world: Nietzsche's texts lend themselves to two possible interpretations—phenomenological and ontological. Before offering in conclusion my own, non-literal reading of the error theory, I shall now provide a brief *reductio ad absurdum* of these two possibilities by showing that neither of them can be accepted without committing Nietzsche to gross inconsistencies.

The first, phenomenological, candidate is what Nietzsche calls the 'fuzziness and chaos of sense impressions' (WP, §569): thus 'there is in us a power to order, simplify, falsify, artificially distinguish. "Truth" is the will to be master over the multiplicity of sensations' (WP, §517).³⁸ As indicated, the idea is that we can only perceive a world which has been ordered into self-identical, causally related spatio-temporal objects. This 'world of "phenomena"' is the adapted world which we feel to be real' (WP, §569). But such reality is merely an illusion. By contrast, 'the antithesis of this phenomenal world is not the "true world" but the *formless, unformulable world of the chaos of sensations*' (WP, §569, my emphasis). This interpretation of the error theory does not rest on metaphysical assumptions about a mind-independent reality to which our perception would fail to conform. It seems to be a new version of the idea already present in *TL*, according to which the use of concepts is *per se* falsificatory. In *TL*, the view was that concepts result from the equalization of differences between various, non-conceptually formed, perceptual metaphors. Here, Nietzsche offers a reverse claim: the *a priori* use of concepts simplifies the primal stream of inchoate sense data which underlies conscious perception. Perhaps the best way to make sense of this idea is to think of it in Husserlian terms, and to suggest that such a primal stream of impressions could be understood as the 'hyle' of perception, 'sensation contents' (*Ideen I*, §85) which are associated through temporal and spatial syntheses (passive syntheses in *Experience and Judgment*), and actively unified into intentional objects through the 'sense bestowing' acts of the mind. However on Husserl's view such sensory contents are not truly sense data in the sense that they are not objects of awareness (this would require a noetic function); moreover, they are never given *on their own* (since we only apperceive intentional wholes for Husserl). They are a *dependent* aspect of conscious experience, which can only be approximated retrospectively, through introspection, and by means of such technical procedures as the *epoché* and the transcendental reduction. Yet try as we may, such reflective procedures can only give us access to synthetized unities. Thus while there may be such a thing as a hyletic layer, it has no independent existence and cannot be described adequately, only evoked by contrast to higher intentional levels, in an abstract manner. Nietzsche's views seem close to Husserl in the sense that, as we have seen, he rejects the idea that we could be aware of raw sense data passively registering on the mind. Thus 'inner experience enters our consciousness only after it has found a language the individual understands' (WP, §479). As in Husserl, such process has higher level layers: for us to experience something, 'it must be experienced

³⁸ Such a phenomenological reading is evoked e.g. by Granier 1977, 137; Wilcox 1974, 133, 149; and Stack 1991, 35.

as adapted, as “recognisable”. The material of the senses adapted by the understanding, reduced to rough outlines, made similar’ (*WP*, §569). But there are also lower layers, which seem closer to the Husserlian passive synthesis of homogeneity: ‘there could be no judgment at all if a kind of equalisation were not practiced within sensation’ (*WP*, §532). As in Husserl again (and presumably for the same reason), such a primary layer is inaccessible to consciousness: it is ‘another kind of phenomenal world, a kind “unknowable” for us’ (*WP*, §569). Yet if this account is correct, then it is difficult to understand how the falsification thesis can hold (and Husserl certainly did not draw this conclusion from his account of perception). The reason is that presumably something can only be falsified if it can be said to exist independently, and to have identifiable intrinsic characteristics which are later distorted in such a way that through comparison with the original, we could pinpoint the falsification. Yet neither on Nietzsche’s account nor on Husserl’s do the primary layers of perception exist on their own; moreover, given the way our minds work it would be impossible for us to know what they are, and therefore to judge whether and how they would have been falsified.

The second, ontological, candidate for what knowledge allegedly falsifies is the ‘world of becoming’.³⁹ According to Nietzsche, such a world ‘could not, in a strict sense, be “comprehended” or “apprehended” or “known”; only to the extent that the “comprehending” and “knowing” intellect encounters a coarse, already created world, fabricated out of mere appearances but become firm to the extent that this kind of appearance has preserved life, only to this extent is there “knowledge”’ (*WP*, §520). This second version of the falsification thesis both differs from and complements the previous one. It has in common the idea that our experience presents us with a world of idealities, the stability of which is made necessary by our practical needs: thus ‘our needs have made our senses so precise that the “same apparent world” always reappears and has thus acquired a semblance of reality’ (*WP*, §521). The new element is that, this time, the object of falsification is not immanent to experience anymore, in the way the ‘chaos of sensations’ was supposed to be. What is falsified is a mind-independent world, a Heraclitean world of perpetual flux, which changes so fast that nothing can even be said to be identical to itself:⁴⁰ ‘the total character of the world, however, is in all eternity chaos’ (*GS* iii, §109). So the idea is now that there is a fundamental incompatibility between our sensory and conceptual apparatus on the one hand, and the ‘real’ world on the other. Thus ‘logic too depends on presuppositions to which nothing in the real world corresponds, for example on the presupposition that there are identical things, that the same thing is identical at different points in time’⁴¹ (*HH* i, §11).

³⁹ For ontological readings of the ‘world of becoming,’ see Danto 1965, 89, 96–7; Grimm 1977, 30, 32; Magnus 1978, 25, 169.

⁴⁰ One may argue, along Kantian lines, that on such a construal it would be impossible for us to notice any changes, or even to have the concept of change. For unless *something* can be said to change, we would be unable to identify any modifications in the flux of becoming, and thus to even have a notion of becoming.

⁴¹ See also: ‘without a constant falsification of the world by means of numbers, man could not live’ (*BGE* i, §4), or again ‘the will to logical truth can only be carried through only after a fundamental falsification of

The result of such falsification is an imaginary world of spatio-temporal objects, causes, and effects, which we wrongly believe to be real: “‘reality’ is always only a simplification for practical ends, or a deception through the coarseness of organs” (*WP*, §580). However, the problem with this reading is that its intrinsic dualism (which contrasts the world of stable appearances with the ‘real’ world of becoming) seems to rest precisely on the sort of ontological commitments rejected by Nietzsche. It is difficult to see how such a world, which supposedly has an independent existence and can only appear to us through the simplifications of our minds, and is thus deemed by definition ‘unknowable’, is different from the in-itself criticized by Nietzsche. On this interpretation then, later Nietzsche would end up in a worse position than his younger self, since at the time of *TL* he refrained from making any assumptions about the nature of the ‘mysterious X’, and pointed out against Schopenhauer that none of the concepts or categories used to structure empirical experience could be applied to it. Yet if the ‘real’ world is identified with the world of becoming (i.e. a concept of empirical origin), Nietzsche would be making both these mistakes, which seems strange given the care he has taken to move away from the ontological assumptions of transcendental idealism. The only way out of this dilemma is to move away from metaphysical commitments and to understand the world of becoming as a mind-independent *empirical* world. Thus Cox claims that the world of becoming is ‘the world we inhabit and know: the natural, physical world, the world of “appearance”’ (Cox 1999, 184). He asserts that this world is ‘empirically evident’ (188), that it is the ‘physical, natural world that we inhabit and with which we are familiar’ (193) and which must be understood with reference to Heraclitus. Yet this option seems very counter-intuitive, not to say implausible: one would think that the very point of Nietzsche’s claims about the world of becoming being chaotic and incomprehensible is to bring out the fact that there is nothing ‘empirically evident’ or familiar about it. What is empirically evident is precisely its opposite, i.e. the world of stable objects we encounter in experience, and which Nietzsche contrasts with the ‘real’ world.

IV

Thus the literal reading of the error theory, whether construed phenomenologically or ontologically, is met with insurmountable difficulties. Yet perhaps something valuable can be gleaned from this failure: it may be taken as an indication that the problem, in this case, lies with literalness itself. I shall therefore turn to a non-literal construal which, I should hasten to say, is intended as a tentative answer only: the

all events is assumed’ (*WP*, §512). Some interesting insights about Nietzsche’s views on logic can be found in Hales and Welshon 2000, 43ff. The authors argue that while syntactic logic, which provides the rules for the manipulation of the operators, connectives, quantifiers, etc. of a formal system, does not need such presuppositions, semantic logic, which specifies domains of entities, does do so. Thus ‘in order for the symbols and formulas of logic to mean anything or have any applicability, there must be sets of things to which they refer’ (Hales and Welshon 2000, 43).

incompatibility of some of the passages cited removes the possibility of a univocal solution to the epistemological quandary resulting from the error theory. My hypothesis is that the real target of Nietzsche's skepticism about knowledge may not be so much the possibility of limited objectivity resulting from the use of perspectival conditions as two kinds of illusions: on the one hand, our inclination, motivated by the drive for survival, to ignore the fact that our experience is determined by transcendental features, which results in metaphysical realism or in naive empirical realism; on the other, the tendency of some philosophers (in particular Kant) to take what are merely features for constraints, and to think that universal and necessary knowledge is possible, when in fact only relative forms of objectivity are legitimate. To use the vocabulary of the quotation already given, the problem may not be so much with our having a grammar, or with the world we experience according to it, as with our lack of awareness of such grammar and our blind 'faith' in the mind-independent reality of the entities that surround us and which we instinctively 'feel to be real' (*WP*, §569). In that case, we should not read the 'world of becoming' literally, as an expression describing the true state of the world, but metaphorically, as pointing towards an irrepresentable object. The function of such an object would not be to serve as a problematic referent for the falsification thesis, but to make us reflect both on the workings of our own minds and on our assumptions about the empirical world: the irrepresentability of the 'world of becoming' draws our attention (a contrario) to the fact that both perception and the perceived objects are structured by some naturalized transcendental features, without which no representation is possible. By pointing towards the limits of human representation, it allows its structural conditions to emerge for reflection.

This suggestion has the advantage of minimizing the self-reference problem and of complementing the main strands of Nietzsche's thought previously identified: his qualified endorsement of transcendental determination, his naturalization of transcendental constraints into transcendental features, and his deliberate agnosticism about ontological commitments. It is clearly compatible with, and helps make sense of, the weaker formulations of the error theory; as for the stronger formulations, they can more fruitfully be regarded as a deliberately hyperbolic rhetorical device meant to signal that what is problematic is not so much the existence of such transcendental features (after all, they are just the expression of our finitude), but the fact that we have a natural tendency to be unaware of their existence. We tend to behave either as metaphysical realists, i.e. as if our knowledge were unconditioned and we had direct access to the in-itself, or as naive commonsense realists, i.e. as if we could just encounter pre-existing, mind-independent objects in the world, without realizing that there is an *a priori* (in Nietzsche's revised sense) agreement between the structure of such objects and the conditions under which we know them, and that therefore such objects are constituted, not found. In other words, the real target of the error theory would not be the possibility of (limited) knowledge itself, but a certain naivety about its scope and its objects. This warning is clearly expressed in the following passage: 'our knowledge is

no knowledge in-itself . . . It is our own laws that we project in the world—even though appearance teaches us the opposite and seems to point towards us as consequences of this world, and to point towards these laws in their action upon us' (KSA, M III 1. Spring–autumn 1881 15[9], my emphasis). Given the ineluctability and the strength of the 'teaching of appearances,' we need the stronger formulations of the error theory to help us resist our natural tendency to believe in the pre-existence and independence both of the empirical world (of which we appear to be 'consequences') and of its laws. What we must beware of is not so much the conditioned character of our knowledge as our ignorance of the 'poetico-logical power' from which it results, and by 'virtue of which we keep ourselves alive' (KSA III 1. Spring–autumn 1881 15[9]): 'it is *we* alone who have devised cause, sequence . . . and when we project and mix this symbol world into things as if it existed "in-itself", we act once more as we always have acted—*mythologically*' (BGE i, §21, Nietzsche's emphasis). On the reading I suggest, the error theory points out that the problem is not so much the projection, since it is constitutive of the way human experience works, as our lack of awareness of our own activity, and our corresponding belief in the independent existence of the projected.

If this is right, then it becomes possible to interpret Nietzsche's views on knowledge in a way which does not make them self-undermining, but on the contrary brings out two interesting aspects: first, the intertwining of naturalistic and transcendental aspects in his account of the formation of human experience, which culminates in the notion of naturalized a priori conditions (transcendental features) and makes it impossible (contra both Green on the one hand, and Clark, Cox, and Leiter on the other) to define him either as a transcendental philosopher or as a naturalist. In this, Nietzsche anticipates later Continental philosophers such as Foucault, who share both his anti-empiricist assumptions about the constitution of experience and his skepticism about the possibility of truly necessary, universal a priori constraints. The Foucauldian notion of a historical a priori, rooted in and yet distinct from contingent historical practices, presents a similar ambiguity (and interest) as Nietzsche's own naturalization of the transcendental, in that both its necessity and its scope are dependent on practices to which it is, however, not reducible, and which presuppose that its existence be intelligible. The second important aspect resides in the highly *critical* dimension of Nietzsche's thought, which manifests itself at two distinct levels: ontologically speaking, through his rejection of the formerly endorsed robust, two-world version of transcendental idealism in favour of a minimalist, perspectivist ontology according to which it makes no sense to talk about things in themselves, let alone to attribute them a self-standing essence. All that can be said is that, should our own perspectival conditions be bracketed, it is possible that others would apply, according to which different, incommensurable worlds would be constituted. But what they apply to could not be defined independently from these other perspectives, and thus could not be said to have any essence in-itself: reality is perspectival through and through. Moreover, while this possibility has to be acknowledged, one must be more cautious than Kant himself was and remain agnostic about the actual existence of such conditions and

worlds. In Kantian terms, the critical impact of Nietzsche's thought here is that the status of the ontological commitments of transcendental idealism must be considered as problematic, not assertoric (let alone apodictic). Epistemologically speaking, this critical dimension is expressed by the role played by his so-called error theory. As we have seen, part of the thrust of his naturalization of the transcendental is to emphasize the fact that *a priori* conditions do not have any rational validation, but are entailed by practices which are necessary for the survival of our species. Thus in order to preserve and develop their lives, human beings have to believe in the existence of an ordered, relatively stable world of mind-independent entities, with which they can interact causally, and on which they can act, obtaining results, with a sufficient degree of predictability. Correlatively, we instinctively tend to view knowledge as the non-perspectival identification of the properties of such pre-existing entities. On this background, I have argued that the function of the error theory is not to suggest that there is something wrong with perception, or that the perceived world is fictitious, but to draw our attention to the dependency of both on naturalized transcendental conditions. It is not so much an error theory *sensu stricto* as a hyperbolic warning against uncritical forms of realism. It is not aimed against the possibility of a statement being true within a set of perspectival conditions, but rather at a set of implicit beliefs: that there are no such conditions (naive realism), or that our statements can be true of all possible perspectives (which would require the existence of universal transcendental constraints), or worse, extra-perspectively (by virtue of metaphysical correspondence). Thus the error theory is meant to counterbalance our hard-wired tendency to be naive realists about the world and about knowledge.⁴² While such a tendency cannot be eradicated (precisely for the sort of naturalistic reasons pointed out), it is possible to limit its effects in two ways: on the one hand, by exposing it, although in order to do so one will have to constantly fight against our nature—hence the deliberately hyperbolic character of the error theory; on the other, by openly indexing one's own statements to their conditions of possibility (hence Nietzsche's constant emphasis on the perspectival aspect of his claims). Thus what makes Nietzsche's views on knowledge particularly interesting, in my view, is that while they are clearly very strongly influenced both by transcendental and naturalistic concerns, they present a position which tends to

⁴² Thus it is meant to deter us from adopting precisely the sort of 'common sense realism' advocated by both Clark and Leiter. In this regard, the thrust of Nietzsche's argument is to suggest that such realism is far from being devoid of metaphysical assumptions about the nature of reality. It implicitly relies on the claim that empirical objects are mind-independent and have intrinsic properties which can be defined independently from the way we access them. Leiter's gloss of the well-known passage about perspectivism in *GM* iii, §12, is illustrative of this belief: 'consider a useful analogy. If we wanted to get knowledge of a particular geographic area by making a map of it, the kind of map we would make would necessarily be determined by our interests. . . . The map corresponding to each set of interests would give us genuine knowledge of the area, and the more interests embodied in maps, the more we would know about the area' (Leiter 2002, 273–4). The striking thing about the analogy is that it presupposes that the area to be charted by the various perspectival maps (i.e. the real) pre-exists and is independent from the charting itself, which is exactly the sort of naive belief Nietzsche warns us against by means of the error theory.

overcome the opposition between the two trends while taking up one of the most valuable insights of the Kantian legacy, namely its critical spirit.

References

Works by Nietzsche

- BGE: *Beyond Good and Evil*, tr. Walter Kaufmann (New York: Vintage Books, 1989).
 GM: *On the Genealogy of Morals*, tr. Walter Kaufman (New York: Vintage Books, 1967).
 GS: *The Gay Science*, tr. Walter Kaufmann (New York: Vintage Books, 1974).
 HH: *Human, All Too Human*, tr. R. J. Hollingdale (Cambridge: Cambridge University Press, 1968).
 KSA: *Kritische Studienausgabe*, ed. Giorgio Colli and Mazzino Montinari (Berlin: Walter de Gruyter, 1980).
 PT: 'The Philosopher', in *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s*, ed. Daniel Breazeale (London: Humanities Press, 1979).
 TI: *Twilight of the Idols*, tr. Richard Polt (Indianapolis: Hackett, 1997).
 TL: 'Truth and Lies', in *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s*, ed. Daniel Breazeale (London: Humanities Press, 1979).
 WP: *The Will to Power*, tr. Walter Kaufmann (New York: Vintage Books, 1968).

Other references

- Allison, Henry (1983), *Kant's Transcendental Idealism: An Interpretation and Defense* (New Haven: Yale University Press).
 Bird, Graham (1962), *Kant's Theory of Knowledge* (London: Routledge).
 Clark, Maudemarie (1990), *Nietzsche on Truth and Philosophy* (Cambridge: Cambridge University Press).
 Clark, Maudemarie (2001), 'The Development of Nietzsche's Later Position on Truth', in John Richardson and Brian Leiter (eds), *Nietzsche* (Oxford: Oxford University Press), 59–84.
 Cox, Christoph (1999), *Nietzsche: Naturalism and Interpretation* (Berkeley, CA: University of California Press).
 Danto, Arthur (1965), *Nietzsche as a Philosopher* (New York: Macmillan).
 De Man, Paul, *Allegories of Reading: Figural Language in Rousseau, Nietzsche, Rilke, and Proust* (New Haven: Yale University Press).
 Derrida, Jacques (1979), *Spurs: Nietzsche's Styles*, tr. Barbara Harlow (Chicago: University of Chicago Press).
 Granier, Jean (1977), 'Nietzsche's Conception of Chaos', in David Allison (ed.), *The New Nietzsche* (Cambridge, MA: MIT Press), 135–41.
 Green, Michael (2002), *Nietzsche and the Transcendental Tradition* (Urbana, IL, and Chicago: University of Illinois Press).
 Grimm, Rüdiger (1977), *Nietzsche's Theory of Knowledge* (Berlin: Walter de Gruyter).
 Hales, Steven, and Rex Welshon (2000), *Nietzsche's Perspectivism* (Chicago: University of Illinois Press).
 Kofman, Sarah (1993), *Nietzsche and Metaphor*, tr. Duncan Large (London: Athlone Press).

- Leiter, Brian (1994), 'Perspectivism in Nietzsche's *Genealogy of Morals*', in Richard Schacht (ed.), *Nietzsche, Genealogy, Morality* (Berkeley, CA: University of California Press), 334–57.
- Leiter, Brian (2002), *Nietzsche on Morality* (London: Routledge).
- Magnus, Bernd (1978), *Nietzsche's Existential Imperative* (Indianapolis: Indiana University Press).
- Nehamas, Alexander (1985), *Nietzsche: Life as Literature* (Cambridge, MA: Harvard University Press).
- Poellner, Peter (1995), *Nietzsche and Metaphysics* (Oxford: Clarendon Press).
- Poellner, Peter (2001), 'Perspectival Truth', in John Richardson and Brian Leiter (eds), *Nietzsche* (Oxford: Oxford University Press), 85–117.
- Richardson John (2004), *Nietzsche's New Darwinism* (Oxford: Oxford University Press).
- Sacks, Mark (2000), *Objectivity and Insight* (Oxford: Oxford Clarendon Press).
- Stack, George (1983), *Lange and Nietzsche* (Berlin and New York: Walter de Gruyter).
- Stack, George (1991), 'Kant, Lange and Nietzsche: Critique of Knowledge', in Keith Ansell-Pearson (ed.), *Nietzsche and Modern German Thought* (London: Routledge), 30–58.
- Strawson, Peter (1966), *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason* (London: Methuen).
- Wilcox, John (1974), *Truth and Value in Nietzsche: A Study of his Metaethics and Epistemology* (Ann Arbor: University of Michigan Press).

10

Husserl and the Transcendental

Dan Zahavi

Do we still need to read Husserl or is the philosophical vision and program he offers us so outdated that the most charitable course of action is to bury it in silence? It is both intriguing and also somewhat odd that some of the most dismissive attacks on Husserl have been made by other phenomenologists, in particular by recent Merleau-Pontians and Heideggerians. Take for instance Philip Dwyer who writes that, although Merleau-Ponty occasionally tried to make excuses for Husserl and even distorted his doctrine in order to make it more palatable, the fact remains that for the most part, Husserl's work was antithetical to Merleau-Ponty's (Dwyer 1990, 33–4), and as he then concludes: 'In my view, what, for the most part, Husserl meant by and practiced as "phenomenology" can only be described as giving new meaning to the word "muddled". The less said about the details of Husserl's philosophy the better' (Dwyer 1990, 34). As for a Heideggerian equivalent, consider Taylor Carman who in a recent book dwells on Heidegger's relation to Husserl and quotes from two letters that Heidegger wrote to Karl Löwith in 1923. In the first of the letters, dating from 20 February 1923, Heidegger writes:

In the final hours of the seminar, I publicly burned and destroyed the *Ideas* to such an extent that I dare say that the essential foundations for the whole [of my work] are now cleanly laid out. Looking back from this vantage to the *Logical Investigations*, I am now convinced that Husserl was never a philosopher, not even for one second in his life. He becomes ever more ludicrous. (Quoted in Carman 2003, 61)

On 8 May 1923, Heidegger again wrote to Löwith, this time to say that he in his lecture course that semester

strikes the main blows against phenomenology. I now stand completely on my own feet. . . . There is no chance of getting an appointment [with Husserl's help]. And after I have published, my prospects will be finished. The old man will then realize that I am wringing his neck—and then the question of succeeding him is out. But I can't help myself. (Quoted in Carman 2003, 62)

Carman considers the substance of both letters—i.e. Heidegger's appraisal of Husserl's philosophical status—to be quite accurate. Carman claims that Husserl's and Heidegger's philosophy are worlds apart in both style and substance and that their respective aims and aspirations are profoundly different (Carman 2003, 54). Indeed, according to Carman, Heidegger has shown

that Husserl's phenomenology is at once uncritical and incoherent: uncritical in its appropriation of the Cartesian conception of the subject and the Platonic-Aristotelian interpretation of being as presence (*Anwesen*); incoherent because it purports to ground those prejudices in a rigorous philosophical method that can itself be made intelligible only by taking them for granted. Husserl's project is thus caught in a vicious circle, for its results presuppose its methods and its methods presuppose its results. Neither genuinely radical nor free of substantive presuppositions, Husserl's phenomenology is simply not the 'rigorous science' it claims to be. (Carman 2003, 54)

Thus, on Carman's reading, the hermeneutic phenomenology of *Being and Time* amounts to a wholesale rejection of Husserl's transcendental phenomenology, with all its Platonism, mentalism, and methodological solipsism (Carman 2003, 56, 62). It is consequently not surprising that Carman also claims that Heidegger rejects Husserl's transcendental reduction since he takes it to represent a failure to acknowledge the hermeneutic constraints on any adequate understanding of human being as being-in-the-world (Carman 2003, 56). But how then does Carman interpret the phenomenological reduction? Here is what he says:

The transcendental reduction . . . consists in methodically turning away from everything external to consciousness and focusing instead on what is internal to it. The reduction thus amounts to a special kind of reflection in which the ordinary objects of our intentional attitudes drop out of sight, while the immanent contents of those attitudes become the new objects of our attention. (Carman 2003, 80)

Indeed, for Husserl, 'Intentionality is internal, the world is external, and the transcendental reduction focuses on the former to the exclusion of the latter' (Carman 2003, 86).

A noticeable feature of both interpretations is that they ignore not only much recent Husserl scholarship but also fail to consider some of Husserl's major works, including not only *Zur Phänomenologie des inneren Zeitbewusstseins* or *Zur Phänomenologie der Intersubjektivität I–III* but also all of the volumes in the Husserliana edition that have been published during the last thirty years. By contrast, consider Merleau-Ponty who was well aware of the importance of Husserl's research manuscripts (which to a large extent are the ones that have been published in the recent Husserliana volumes). As he wrote in a letter dating from 1942: 'After all, Husserl's philosophy is almost wholly contained in his unpublished manuscripts' (quoted in Van Breda 1962, 420). This was an appraisal that matched Husserl's own view. As the latter wrote in a letter to Adolf Grimme in 1931: 'Indeed, the largest and, as I actually believe, most important part of my life's work still lies in my manuscripts, scarcely manageable because of their volume' (Husserl 1973c, lxvi). Moreover, as Husserl was at pains to point out, many

of his critics had been more concerned with what his former students had to say about his philosophy than with the study of his own writings:

From the start one thinks one knows what it is all about . . . In the best case one has read my writings; more often, one has sought advice from my students who, having been taught by me, are supposedly able to provide reliable information. Thus one is guided by the interpretations of Scheler, Heidegger, and others, and thereby spares oneself the admittedly very difficult study of my own writings. Protests against this are met with a ready answer: the old man is stuck in his established train of thought and is unreceptive to any refuting criticism . . . Here at the outset I require only this one thing, that one keep these sorts of prejudice, one's knowing in advance the meaning of those words that I have furnished with entirely new sense: phenomenology, transcendental, idealism . . . firmly locked away in one's breast . . . Initially, one hears and sees what is being presented, one goes along and sees where it might lead and what might be accomplished with it. (Husserl 1954, 439–40)

In order to get a sound grasp of the basic nature of Husserl's phenomenology, it is crucially important to recognize that, although Husserl used many traditional terms, the use of these terms was by no means traditional. This can be exemplified in a number of ways (see Zahavi 1994, 1999, 2002a, 2002b, 2008) but let me in the following focus on Husserl's notion of the transcendental. This focus will necessitate some reflections on Husserl's methodology and on his relation to Kant. And to anticipate my conclusion, I will basically argue that, although Husserlian phenomenology certainly deserves to be classified as a form of transcendental philosophy, it also differs rather markedly from traditional forms of transcendental philosophy.

1. The Transcendental Reduction

In 1925 Husserl wrote a letter to Ernst Cassirer where he described the development of his own appreciation of Kant in some detail (Husserl 1994, 5/4). Initially, Husserl had been strongly influenced by Brentano's negative appraisal of Kant, but subsequent studies made Husserl realize the affinity between his own project and that of Kant. There is, of course, one place where Kant's influence on Husserl is particularly visible. As Husserl admits in *Erste Philosophie I*, when he decided to designate his own phenomenology as transcendental, he was exactly making use of a Kantian concept (Husserl 1956, 230).

One way to interpret Kant's revolutionary Copernican turn is by seeing it as amounting to the realization that our cognitive apprehension of reality is more than a mere mirroring of a pre-existing world. Moreover, transcendental philosophy transformed the pre-critical search for the most fundamental building blocks of empirical reality into a reflection on what conditions something must satisfy in order to count as 'real'; what is the condition of possibility for the appearance of empirical objects?

Why does Husserl's phenomenology merit the name transcendental? Husserl's standard answer is that phenomenology is transcendental because its aim is to

clarify the constitution of transcendence (Husserl 1974, 259). Or as he puts it in *Cartesianische Meditationen*, the two concepts transcendence and transcendental are correlated and the task of transcendental phenomenology consists in *elucidating* mundane transcendence through a systematic disclosure of constituting intentionality (Husserl 1950, 34, 65). What does this amount to? Husserl concedes that traditional epistemology has also been confronted with the problem of transcendence, but in its traditional form, the problem has been how certainties and evidences pertaining to the immanency of conscious life can gain objective significance (Husserl 1950, 116), or to put it more bluntly, the problem has been how to get outside the sphere of consciousness. But as Husserl makes clear, this rendering of the problem makes it completely nonsensical. It presents us with a pseudo-problem, which only arises if one forgets the true lesson of intentionality and conceives of the mind as an isolated world-detached entity. In truth, it is a decisive mistake to conceive of knowledge as a relation between two independently variable dimensions, as if mind and world only by coincidence happened to fit each other (Husserl 2003, 30). For Husserl, mind and world are not distinct entities; rather they are interdependent, they are bound constitutively together.

Husserl frequently argues that his transcendental phenomenology can be seen as an attempt to redeem rather than renounce the realism of the natural attitude. In fact, Husserl would claim that a transcendental reflection or methodology is required if we are to understand and account for the realism that is intrinsic to the natural attitude. This is why Husserl can write that his *transcendental idealism* contains natural realism within itself (Husserl 1962, 254), since it is an explication of the sense that the world has for all of us 'prior to any philosophizing' (Husserl 1950, 36). This last statement fits well with Husserl's repeated insistence that the aim of his investigations is to understand and clarify the legitimacy of our belief in reality, it is not to critically evaluate or justify it (since it is not in need of any such justification):

[T]he transcendent world; human beings; their intercourse with one another, and with me, as human beings; their experiencing, thinking, doing, and making, with one another: these are not annulled by my phenomenological reflection, not devalued, not altered, but only understood. (Husserl 1974, 282)

There can be no stronger realism than this, if by this word nothing more is meant than: 'I am certain of being a human being who lives in this world, etc., and I doubt it not in the least.' But the great problem is precisely to understand what is here so 'obvious'. (Husserl 1954, 190–1)

That the world exists, that it is given as existing universe in uninterrupted experience which is constantly fusing into universal concordance, is entirely beyond doubt. But it is quite another matter to understand this indubitability which sustains life and positive science and to clarify the ground of its legitimacy. (Husserl 1971, 152–3)

It is against the background of such considerations that Husserl explicitly denies that the performance of the (in)famous transcendental reduction entails any loss, except of course, the loss of naivety. The true purpose of the *epoché* and the

reduction is not to doubt, neglect, abandon, or exclude reality from our research; rather their aim is to suspend or neutralize a certain dogmatic *attitude* toward reality, thereby allowing us to focus more narrowly and directly on reality just as it is given. To put it differently, the *epoché* and the reduction do not involve an exclusive turn toward inwardness. On the contrary, they permit us to investigate reality in a new way, namely in its significance and manifestation for consciousness. In short, they entail a change of attitude toward reality, and not an exclusion of reality. As Husserl points out in the lecture *Phänomenologie und Anthropologie* which he gave in Frankfurt, Berlin, and Halle in 1931, the only thing that is excluded as a result of the *epoché* is a certain naivety, the naivety of simply taking the world for granted, thereby ignoring the contribution of consciousness (Husserl 1989, 173). And as Husserl repeatedly insists in this 1931 lecture, the turn from a naive exploration of the world to a reflective exploration of the field of consciousness does not entail a turning away from the world, rather, it is a turn that for the first time allows for a truly radical investigation and comprehension of the world (Husserl 1989, 178). Such a reflective investigation obviously differs from a straightforward exploration of the world, but it remains an investigation of reality; it is not an investigation of some otherworldly, mental, realm. Only a mistaken view of the nature of meaning and appearance would lead to such a misunderstanding. Thus, as Husserl has repeatedly emphasized in many of his writings, to perform the *epoché* and the reduction makes a decisive discovery possible and should consequently be understood as an *expansion* of our field of research (Husserl 1950, 66; 1954, 154). This is why Husserl in *Krisis* can compare the performance of the *epoché* with the transition from a two-dimensional to a three-dimensional life (Husserl 1954, 120). In fact, as he points out in *Erste Philosophie II*, it is better to avoid using the term ‘Ausschaltung’ altogether, since the use of this term might easily lead to the mistaken view that the being of the world is no longer a phenomenological theme, whereas the truth is that transcendental research includes ‘the world itself, with all its true being’ (Husserl 1959, 432).

The interpretation just offered is further supported by material found in the recently published Husserliana 34, which carries the title *Zur phänomenologischen Reduktion: Texte aus dem Nachlass*. In these texts, which were written between 1926 and 1935, Husserl makes it clear that when the natural world is spoken of as ‘ausgeschaltet’, what this means is merely that the transcendental philosopher must cease to posit the world naively (Husserl 2002, 21). It doesn’t imply that I cannot continue to observe, thematize, and make judgments concerning the world, but I must do so in a reflective manner that considers the world as intentional correlate (Husserl 2002, 58). To put it differently, to perform the *epoché* and the reduction is to effectuate a thematic reorientation. Henceforth the world is revealed as a phenomenon, and as such it remains in the very focus of my phenomenological research (Husserl 2002, 204, 323). To quote Husserl:

The world as ‘phenomenon’, as world in the *epoché*, is merely a modality, in which the same ego, for whom the world is pregiven, reflects on this pregivenness and on what it contains, and for that reason doesn’t abandon the world and its validity, nor makes it disappear. (Husserl 2002, 223; cf. Husserl 2002, 83–4)

To consider the world as a phenomenon is consequently not an abstraction, but rather a thematization of an otherwise anonymously functioning dimension, namely the very dimension of givenness. As Husserl puts it, to effectuate the reduction is to liberate the world from a hidden abstraction, and to reveal it in its concretion as a constituted network of meaning (*Sinngebilde*) (Husserl 2002, 225). At the same time, the reduction also liberates the I from the limitations of its natural being. My ordinary natural life is according to Husserl a life lived in self-alienation (*Selbstentfremdung*), since it doesn’t know of its transcendentality. The performance of the reduction removes the blinkers (*Scheuklappen*) that ordinarily conceals the full and concrete transcendental character of life (Husserl 2002, 226, 233).

By adopting the phenomenological attitude we pay attention to the givenness of public objects (trees, planets, paintings, symphonies, numbers, states of affairs, social relations, etc.). But we do not simply focus on the objects precisely as they are given; we also focus on the subjective side of consciousness, thereby becoming aware of our subjective accomplishments and of the intentionality that is at play in order for the objects to appear as they do. When we investigate appearing objects, we also disclose ourselves as datives of manifestation, as those to whom objects appear. The topic of the phenomenological analyses is consequently not a worldless subject, and phenomenology does not ignore the world in favor of consciousness. On the contrary, on Husserl’s account, phenomenology is interested in consciousness because it is world-disclosing. Phenomenology should therefore be understood as a philosophical analysis of the different types of givenness (perceptual, imaginative, recollective, etc.), and in connection with this as a reflective investigation of those structures of experience and understanding that permit different types of beings to show themselves as what they are. Husserl would consequently deny that the task of phenomenology is simply to describe objects or experiences as precisely and meticulously as possible. As Scheler once remarked, this would be to make do with a ‘picture-book phenomenology’ (Scheler 1927, vii). Rather, phenomenology is interested in the very dimension of givenness or appearance and seeks to explore its essential structures and conditions of possibility. Such an investigation of the field of presence is beyond any divide between psychical interiority and physical exteriority, since it is an investigation of the dimension in which any object—be it external or internal—manifests itself (cf. Heidegger 1986, 419; Waldenfels 2000, 217).

That Husserl’s phenomenology qualifies as a form of transcendental philosophy shouldn’t really require any extensive argumentation—though whether it resembles or differs from the Kantian version of transcendental philosophy is something in need of further exploration.

2. Husserl, Kant, and the Problem of Intersubjectivity

When it comes to an appraisal of Husserl's phenomenology a reference to Kant can be quite appropriate. This is in particular the case when one is confronted with the claim that Husserlian phenomenology is merely a form of introspective psychology—as Dennett occasionally seems to think (cf. Zahavi 2007). However, when this is said, one should be very careful not to overlook some rather decisive differences between Kant and Husserl. To put it differently, it would be a mistake to think that transcendental philosophy is all one thing, and to overlook the difference between a Kantian transcendental philosophy, and the form of transcendental philosophy we find in phenomenology.

Already in the *Logische Untersuchungen*, Husserl criticized Kant for not having managed to stay clear of a metaphysically contaminated epistemology, and he claimed that metaphysical theories are uncalled for when it comes to an understanding of the relation between the laws of nature and the laws of reason. What are needed are not *explanations*, but phenomenological *clarifications* of meaning, thinking, and knowing (Husserl 1984, 729, 732). Later on, we find Husserl faulting Kant for not having had a proper concept of the a priori, for operating with a too-strong distinction between sensation and understanding, for being too oriented toward the natural sciences, for confusing noetic and noematic analyses, and for lacking methodological rigor (Husserl 1956, 198–9, 235, 282; 1954, 420–1; 1952b, 128; 1976, 246; 1950, 48). When it comes to the difference in method, a telling statement can be found in a manuscript from 1920, where Husserl writes as follows: 'Kant's deduction is a masterpiece of top-down transcendental reasoning. It remains far removed from all phenomenological analyses.' And as he then adds, such a deduction can only be met with disapproval (*Kopfschütteln*) by phenomenologists (quoted in Kern 1964, 104). In one of his longest texts on Kant, *Kant und die Idee der Transzendentalphilosophie*, a text written and presented in commemoration of Kant's bicentennial in 1924, Husserl writes that transcendental philosophy should be based upon a systematic description and analysis of consciousness in all of its modalities (Husserl 1956, 234–5). Husserl distances himself from any sort of regressive transcendental argumentation and criticizes Kant's method for being regressive-constructive. It lacks an intuitive basis and is unable to provide us with a proper account of consciousness. In fact, as Husserl points out in the conclusion, phenomenology insists upon an in-depth investigation of consciousness, and this demand necessitates an *extension* of Kant's concept of the transcendental. It proves necessary to include the humanities and the manifold of human sociality and culture in the transcendental analysis (Husserl 1956, 282). This remark is amplified a few years later when Husserl writes that the possibility of a transcendental elucidation of subjectivity and world is lost if one follows the Kantian tradition in interpreting transcendental subjectivity as an isolated ego and thereby ignores the problem of transcendental

intersubjectivity (Husserl 1993, 120). Thus, it is no coincidence that Husserl at times described his own project as a *sociological* transcendental philosophy (Husserl 1962, 539). In lectures given in London in 1922, Husserl even declared that the development of phenomenology necessarily implied the step from an ‘‘egological’’ . . . phenomenology into a transcendental sociological phenomenology having reference to a manifest multiplicity of conscious subjects communicating with one another’ (Husserl 1981, 68). Ultimately, and this is something that only recently has been properly appreciated, Husserl’s later phenomenology can be seen as an explicit defense of what might be called an intersubjective transformation of transcendental philosophy (cf. Zahavi 1996a, 1996b).

It has often been argued that intersubjectivity can serve as a litmus test for Husserl’s transcendental phenomenology. I agree, but in my view the right way to proceed is not by presupposing a particular understanding of what transcendental philosophy amounts to in order then to test whether it can deal with the issue of intersubjectivity. Rather, one should do exactly the reverse. Thus, I would claim that the genuine sense of Husserl’s transcendental phenomenology can only be disclosed when his theory of intersubjectivity is taken into account. This reversal also seems to be confirmed by Husserl himself, since he writes that his reflections on intersubjectivity have made the ‘full and proper sense’ of transcendental phenomenology understandable for the first time (Husserl 1950, 176).

The link between the role of intersubjectivity and the status of transcendental philosophy is confirmed again and again by Husserl. He frequently writes that his phenomenological treatment of intersubjectivity has the goal of bringing his constitutive analyses to completion—a completion that is achieved the moment it is realized that transcendental intersubjective sociality is the basis in which all truth and all true being have their intentional source (Husserl 1950, 35, 182; 1959, 449; 1962, 295, 344, 474).

In fact, it is not only the completion, but also the realization of transcendental phenomenology that calls for an investigation of intersubjectivity (Husserl 1950, 69; cf. Husserl 1962, 345; 1959, 464). As we have already seen, phenomenology is transcendental precisely because it deals with the problem of how transcendence is to be understood constitutively. And this clarification is, as Husserl repeatedly states, only possible by way of an analysis of intersubjectivity (Husserl 1959, 465; 1974, 259; 1950, 10). As he puts it in two late texts:

The transcendence of the world consists in its being constituted by means of others, by means of the *generatively constituted co-subjectivity*. It is through them that it acquires its ontic sense (*Seinssinn*) as an infinite world. (Husserl 2006, 393)

Here we have the only transcendence that is genuinely worthy of the name—and everything else that is still called transcendence, such as the objective world, rests on the transcendence of foreign subjectivity. (Husserl 1959, 495)

Husserl also describes the other as *the intrinsically first other* (Husserl 1950, 137; 1974, 248) and writes—in a polemic remark directed at Scheler—that only constitutive

phenomenology has recognized the true sense and true scope of the problem of other minds, namely, by recognizing how ‘the otherness of “someone else” becomes extended to the whole world, as its “Objectivity”, giving it this sense in the first place’ (Husserl 1950, 173).

Elsewhere Husserl states that the reduction to transcendental subjectivity simultaneously means reduction to the transcendental intersubjectivity made accessible within it (Husserl 1973c, 73–5, 403). Or as he makes clear in *Zur Phänomenologie der Intersubjektivität III*, the introduction of the intersubjective dimension does not mean some kind of external expansion of transcendental subjectivity, rather it simply expresses a better understanding of what subjectivity amounts to in the first place (Husserl 1973c, 17). As he continues later in the same volume:

I have to distinguish: the currently transcendental-phenomenologizing subjectivity (as real *ego*-monad) and the transcendental subjectivity as such; the latter turns out to be the transcendental intersubjectivity, which contains the transcendental phenomenologizing within itself. (Husserl 1973c, 74–5)

This idea is one that can be found quite frequently in Husserl. In *Erste Philosophie II*, for instance, Husserl writes that transcendental subjectivity in its full universality is exactly *intersubjectivity* (Husserl 1959, 480); in a research manuscript from 1927, which has been published in *Zur Phänomenologie der Intersubjektivität I*, Husserl writes that the absolute reveals itself as the intersubjective relation between subjects (Husserl 1973a, 480). He also claims that the subject can only be world-experiencing insofar as it is a member of a community (Husserl 1950, 166) and that the ego is only what it is as a *socius*, that is, as a member of a sociality (Husserl 1973c, 193).

Reflections like these ultimately led Husserl to the conclusion that absolute being—the being of the transcendental subject—is being as a part of transcendental intersubjectivity. Or as he eventually put it in a famous quote from *Krisis* that Merleau-Ponty was later to discuss in detail: ‘subjectivity is what it is—an ego functioning constitutively—only within intersubjectivity’ (Husserl 1954, 175).

It is difficult to exaggerate the far-reaching implications of this. I think that Stephan Strasser was at least partly right when he back in 1975 wrote that the publication of Husserliana 13–15, i.e. Husserl’s research manuscripts on intersubjectivity, had uncovered material that made all current views about the content of Husserl’s philosophy inadequate (Strasser 1975, 33). Let me try to explain:

If one accepts Husserl’s conviction that reality is intersubjectively valid and that my reality-positing acts are dependent upon my experience of others, one is bound to take not only the consensus but also the *dissent* of the world-experiencing subjects seriously. Husserl’s extended analyses of this problem eventually made him enter fields that have traditionally been reserved for psychopathology, sociology, anthropology, and ethnology. Whereas a strict Kantian transcendental philosophy would have considered such empirical and mundane domains as without any transcendental relevance, due to his interest in transcendental intersubjectivity, Husserl was forced to consider these from

a transcendental point of view (cf. Husserl 1973c, 391). Thus, I believe that Husserl's late thinking is characterized by a decisive expansion of the transcendental sphere; an expansion which was brought about by his interest in intersubjectivity, and which ultimately forced him to consider the transcendental significance of generativity, tradition, historicity, and normality.

Let me focus on the problem of *normality*, which Husserl dealt with intensively in different contexts, and which he considers a constitutive core-concept. Basically, Husserl claims that our experiences are guided by anticipations of normality. We apprehend, experience, and constitute in accordance with the normal and typical structures, models and patterns which our earlier experiences have sedimented in our mind (Husserl 1966, 186). If what we experience happens to clash with our earlier experiences—if it is different—we have an experience of *anormality*, which subsequently leads to a modification of our anticipations (Husserl 1973c, 438). Originally Husserl examined this process in connection with his analysis of the *passive synthesis*, but it is not only at work in the solitary subject. As Husserl says, I have been together with people as long as I remember, and my anticipations are therefore structured in accordance with the intersubjectively handed-down forms of apperception (cf. Husserl 1973b, 117, 125; 1973c, 136). Normality is also *conventionality*, which in its being transcends the individual (Husserl 1973c, 611). Already in *Ideen II* Husserl pointed to the fact that there, next to the tendencies originating from other persons, also exist indeterminate general demands made by custom and tradition: 'One' judges thus, 'one' holds the fork in such and such a way, etc. (Husserl 1952a, 269). What is normal I learn from others (and first and foremost from my closest relatives, that is by the people by whom I am brought up, and who educate me (Husserl 1973c, 428–9, 569, 602–4)), and I am thereby involved in a common tradition, which through a chain of generations stretches back into a dim past.

As I have just mentioned, one consequence of Husserl's treatment of intersubjectivity is that he also had to take the *disagreement* between world-experiencing subjects seriously. If my constitution of objectivity is dependent upon my assurance that others experience or can experience the same as I, it is a problem if they claim to be experiencing something different—although the fact that we can agree upon there being a disagreement already indicates a common ground (Husserl 1973c, 47). It is, however, in this context that Husserl emphasizes that only the (dis)agreement between the *normal* members of the community is of relevance. When it is said that real being has to be experientially accessible to everybody, we are dealing with a certain averageness and idealization (Husserl 1973c, 141, 231, 629). 'Everybody' is the person who belongs to a normality of subjects, and who is exactly normal in and through the community (Husserl 1973c, 142). Only with her do we fight over the truth and falsity, being and non-being of our common lifeworld. Only the normal is apprehended as being co-constitutive (Husserl 1973c, 162, 166; 1962, 497), whereas my disagreement with an abnormal is (at first) considered inconsequential.

It is, however, necessary to differentiate between at least two fundamental types of normality. First of all, we speak of normality when we are dealing with a mature, healthy, and rational person. Here the abnormal will be the infant, the blind, or the schizophrenic. Secondly, we speak of normality when it concerns our own *homeworld*, whereas anormality is attributed to the foreigner, which if certain conditions are fulfilled can be apprehended as a member of a *foreign normality*.

It is in this context that the disagreement gains a vital constitutive significance. According to Husserl, the experience of discrepancy between normal subjects (including the experience of a plurality of normalities, each of which has its own notion of what counts as true) does not merely lead to a more complex world-comprehension insofar as we, if we are able to synthesize the standpoints, can gain a *richer insight*. The disagreement can also *motivate* the constitution of *scientific* objectivity, insofar as we aim toward reaching a truth that is valid for us all (Husserl 1954, 324). Thus, eventually it becomes necessary to differentiate between (1) 'normal' objectivity, which is correlated with a limited intersubjectivity (a community of normal subjects) and (2) 'rigorous' objectivity, which is correlated with the unlimited totality of all subjects (Husserl 1973b, 111).

Husserl consequently believed a correlation to exist between different levels of normality and different levels of objectivity (Husserl 1973c, 155). Even absolute objective being and truth is correlated with a subject-dependent normality: the normality of rational subjects (Husserl 1973c, 35–6).

As should be clear from the preceding, however, it is crucial not to interpret Husserl's emphasis on the importance of intersubjectivity as a replacement or substitute for the notion of subjectivity. For Husserl, it only makes sense to speak of intersubjectivity as long as there is a (possible) plurality of subjects, and intersubjectivity can therefore neither precede nor found the individuality and difference of the single subjects. For the very same reason, Husserl would insist that the appeal to intersubjectivity presupposes a commitment to the notion of subjectivity. Intersubjectivity represents an expansion and enrichment of the latter notion; it doesn't constitute an alternative.

Husserl is undoubtedly a transcendental philosopher. But in my view, the kind of transcendental philosophy he is advocating is one that is quite aware of the finitude of the transcendental subject. This is clear not only from Husserl's appeal to a plurality of transcendental subjects, but also from his accentuation of the ongoing and unfinished character of transcendental reflection. In the volume *Zur Phänomenologie der Intersubjektivität III*, Husserl states that the being and truth of absolute objectivity corresponds to a subject-relative normality (cf. Husserl 1973c, 35), and that its constitution can accordingly be understood as the culmination of the development of transcendental intersubjectivity, which is to be conceived precisely as an ongoing unified process of cultivating an ever newer system of norms at ever higher levels (Husserl 1973c, 421). Ever new generations cooperate in transcendently building up the structures of validity pertaining to the objective world, which is precisely a world handed down in tradition (Husserl 1973c, 463). There is, as Husserl writes, no *fixed*

world; rather, it is what it is for us only in the relativity of normalities and abnormalities (Husserl 1973c, 212, 381; 1954, 270). Normality is a tradition-bound set of norms. I learn what counts as normal from others, and I thereby partake in a common tradition. This is why Husserl also refers to *normal life* as *generative life*, and states that every (normal) human being is *historical* by virtue of being constituted as a member of a historically enduring community (Husserl 1973c, 138–9, 431). As he puts it in a manuscript from some time in the 1920s:

That which I have constituted originally (primally instituted) is mine. But I am a ‘child of the times’; I am a member of a we-community in the broadest sense—a community that has its tradition and that, for its part, is connected in a novel manner with the generative subjects, the closest and the most distant ancestors. And these have ‘influenced’ me: I am what I am as an heir. What is really and originally my own? To what extent am I really primally instituting (*urstiftend*)? I am it on the basis of the ‘tradition’; everything of my own is founded, in part through the tradition of my ancestors, in part through the tradition of my contemporaries. (Husserl 1973b, 223)

That Husserl tried to add a historical dimension to transcendental philosophy can also be illustrated in a different way. In a passage quoted earlier, Husserl writes that the transcendence of the world is constituted through others and through the generatively constituted co-subjectivity. Exactly this concept of *generative intersubjectivity* (Husserl 1973c, 199; cf. Steinbock 1995) indicates that Husserl no longer regarded the birth and death of the subject as mere contingent facts, but as transcendental conditions of possibility for the constitution of the world (Husserl 1973c, 171). As Husserl would write in *Krisis*, being embedded in ‘the unitary flow of a historical development’—in a generative nexus of birth and death—belongs as indissolubly to the I as does its temporal form (Husserl 1954, 256).

It is of course debatable whether Husserl’s attempt to implement an intersubjective transformation of transcendental philosophy is a fruitful approach, but it should be evident that Husserl did not advocate a classical Cartesian-Kantian subject-philosophy, and that he was not a solipsist, but on the contrary, treated intersubjectivity as a transcendental philosophical notion of utmost importance.

3. A New Understanding of the Transcendental

Where does this bring us when it comes to an appraisal of Husserl’s notion of the transcendental? One commentator has recently argued that Husserl through the 1920s and 1930s ‘became increasingly wide-reaching, even baroque, in his conception of the transcendental’ (Moran 2002, 51). But rather than calling Husserl’s notion of the transcendental baroque, perhaps it would be more to the point to realize that Husserl subjected the very notion of the transcendental to a far-reaching transformation. As I have argued elsewhere, Husserl’s later phenomenology is characterized by its attempt to modify the static opposition between the transcendental and the mundane, between

the constituting and the constituted (Zahavi 1996b, 2003). It is for instance against this background that one should understand assertions from *Cartesianische Meditationen* to the effect that the constitution of the world implies a mundanization of the constituting subject (Husserl 1950, 130)—that is, that the subject's constitutive experience of the world goes hand in hand with the subject's constitutive experience of its own worldly being. In fact, Husserl ultimately argued that transcendental subjectivity must necessarily conceive of itself as a worldly being if it is to constitute an objective world, since objectivity can only be constituted by a subject which is both *embodied* and *socialized* (Husserl 1950, 130; 1952b, 128). This was not an insight that Husserl only reached at the very end of his life. In a text written around 1914–15, Husserl argues that actual being, or the being of actual reality, doesn't simply entail a relation to some formal cognizing subject, but that the constituting subject in question must necessarily be an embodied and embedded subject. Thus already in this period, Husserl is claiming that the subject in order to constitute the world must necessarily be bodily embedded in the very world that it is seeking to constitute (Husserl 2003, 133). And as he then continues, the constitution of an objective world also requires that the subject stands in an essential relation to an open plurality of other embodied and embedded subjects (Husserl 2003, 135). Such statements fit rather uneasily with the standard assumption that Husserl is a methodological solipsist, an archetypical internalist, and that he should consider the mind some kind of self-sufficient constitutive principle. Husserl's formulations and terminology are not always transparently clear, but I think the view he eventually came to adopt amounts to something like the following. Husserl consistently held the view that reality can only appear thanks to subjectivity. But eventually he came to the realization that the subject does not remain untouched by its constitutive performance, but is, on the contrary, drawn into it, just as constitution is not simply a relation involving a single subject and the world, but rather something that must be considered an inter-subjective process. The problem he then faced was to clarify the exact interrelation between self, world and intersubjectivity. This is made most explicit in his last writings, where the three are increasingly intertwined (cf. Husserl 1959, 505; 1973c, 373).

But let me return to the phenomenological notion of the transcendental. In *Les Mots et les choses*, Foucault has argued that phenomenology exemplifies a type of modern discourse that in its investigation of experience seeks to both separate as well as integrate the empirical and the transcendental. It is an investigation of experience that in the face of positivism has tried to restore the lost dimension of the transcendental, but which at the same time has made experience concrete enough to include both body and culture. To Foucault it is quite clear that this modern type of transcendental reflection differs from the Kantian type by taking its point of departure in the paradox of human existence rather than in the existence of natural science. Although Husserl had apparently succeeded in unifying the Cartesian theme of the cogito with the transcendental motif of Kant, the truth is that Husserl was only able to accomplish this union in so far as he changed the very nature of transcendental analysis. When transcendental subjectivity is placed in the more fundamental

dimension of time, the strict division between the transcendental and the empirical is compromised. The questions of validity and of genesis become entangled. It is this transformation that in Foucault's view has resulted in phenomenology's simultaneously promising and threatening proximity to empirical analyses of man (Foucault 1966, 331–6).

I think Foucault's diagnosis is correct (though I do not share his subsequent criticism of phenomenology). To put it differently, I think one must realize that phenomenological transcendental philosophy differs from Kantian transcendental philosophy. One of the defining features of phenomenological thought has been its attempt to make the coexistence between the transcendental and the empirical perspective less paradoxical. Rather than conceiving of the two as mutually incompatible, they are seen as intertwined and complementary perspectives.

The fact that Husserl's phenomenology operates with an enlarged notion of the transcendental, the fact that it includes topics such as embodiment and intersubjectivity in its transcendental analysis, gives it an advantage in comparison with a more traditional Kantian type of transcendental philosophy. But, of course, it would also be fair to say that this transformation generates new problems of its own. If a transcendental investigation cannot ignore the historicity of human life, if transcendental structures develop over the course of time and can be modified under the influence of experience, it is faced with the task of countering the threat of historical relativism. It should be clear, however, that Husserl, by endorsing the view that the only justification obtainable and the only justification required is one that is internal to the world of experience and to its intersubjective practices, offers a view on the transcendental that points forward in time rather than backwards to Kant. In that sense, and to that extent, Husserl's conception of the transcendental is distinctly modern.¹

References

- Carman, Taylor (2003), *Heidegger's Analytic: Interpretation, Discourse and Authenticity in Being and Time* (Cambridge: Cambridge University Press).
- Dwyer, Philip (1990), *Sense and Subjectivity: A Study of Wittgenstein and Merleau-Ponty* (Leiden: Brill).
- Foucault, Michel (1966), *Les Mots et les choses* (Paris: Gallimard).
- Heidegger, Martin (1986), *Sein und Zeit* (Tübingen: Max Niemeyer).
- Husserl, Edmund (1950), *Cartesianische Meditationen und Pariser Vorträge*, Husserliana 1 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1952), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II*, Husserliana 4 (The Hague: Martinus Nijhoff).

¹ An earlier version of this chapter has appeared in German as 'Phänomenologie und Transzendentalphilosophie', in Günter Figal and Hans-Helmuth Gander (eds), *Heidegger und Husserl: Neue Perspektiven* (Frankfurt am Main: Vittorio Klostermann, 2009), 73–99.

- Husserl, Edmund (1952b), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie III*, Husserliana 5 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1954), *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie. Eine Einleitung in die phänomenologische Philosophie*, Husserliana 6 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1956), *Erste Philosophie I (1923–24)*, Husserliana 7 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1959), *Erste Philosophie (1923/24). Zweiter Teil: Theorie der phänomenologischen Reduktion*, Husserliana 8 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1962), *Phänomenologische Psychologie*, Husserliana 9 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1966), *Analysen zur passiven Synthesis*, Husserliana 11 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1971), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Drittes Buch: Die Phänomenologie und die Fundamente der Wissenschaften*, Husserliana 5 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1973a), *Zur Phänomenologie der Intersubjektivität I*, Husserliana 13 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1973b), *Zur Phänomenologie der Intersubjektivität II*, Husserliana 14 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1973c), *Zur Phänomenologie der Intersubjektivität III*, Husserliana 15 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1974), *Formale und Transzendente Logik*, Husserliana 17 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1976), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie I*, Husserliana 3/1–2 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1981), *Shorter Works*, ed. Peter McCormick and Frederick A. Elliston (Notre Dame, IN: University of Notre Dame Press).
- Husserl, Edmund (1984), *Logische Untersuchungen II*, Husserliana 19/1–2 (The Hague: Martinus Nijhoff).
- Husserl, Edmund (1989), *Aufsätze und Vorträge (1922–1937)*, Husserliana 27 (Dordrecht: Kluwer).
- Husserl, Edmund (1993), *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie. Ergänzungsband. Texte aus dem Nachlass 1934–1937*, Husserliana 29 (Dordrecht: Kluwer).
- Husserl, Edmund (1994), *Briefwechsel—Wissenschaftlerkorrespondenz*, Husserliana Dokumente 3/7, ed. K. Schuhmann (Dordrecht: Kluwer).
- Husserl, Edmund (2002), *Zur phänomenologischen Reduktion. Texte aus dem Nachlass (1926–1935)*, Husserliana 34 (Dordrecht: Kluwer).
- Husserl, Edmund (2003), *Transzendentaler Idealismus. Texte aus dem Nachlass (1908–1921)*, Husserliana 36 (Dordrecht: Kluwer).
- Husserl, Edmund (2006), *Späte Texte über Zeitkonstitution (1929–1934)—Die C-Manuskripte. Materialienband 8* (Dordrecht: Springer).
- Kern, Iso (1964), *Husserl und Kant. Eine Untersuchung über Husserls Verhältnis zu Kant und zum Neukantianismus* (The Hague: Martinus Nijhoff).

- Moran, Dermot (2002), 'Making Sense: Husserl's Phenomenology as Transcendental Idealism', in Jeff Malpas (ed.), *From Kant to Davidson: Philosophy and the Idea of the Transcendental* (London: Routledge), 48–74.
- Scheler, Max (1927), *Der Formalismus in der Ethik und die materiale Wertethik* (Halle: Max Niemeyer).
- Steinbock, Anthony J. (1995), *Home and Beyond: Generative Phenomenology after Husserl* (Evanston, IL: Northwestern University Press).
- Strasser, Stephan (1975), 'Grundgedanken der Sozialontologie Edmund Husserls', *Zeitschrift für philosophische Forschung*, 29, 3–33.
- Van Breda, H. L. (1962) 'Maurice Merleau-Ponty et les Archives-Husserl à Louvain', *Revue de métaphysique et de morale*, 67, 410–30.
- Waldenfels, Bernhard (2000), *Das leibliche Selbst. Vorlesungen zur Phänomenologie des Leibes* (Frankfurt am Main: Suhrkamp).
- Zahavi, Dan (1994), 'Husserl's Phenomenology of the Body', *Études Phénoménologiques*, 19, 63–84.
- Zahavi, Dan (1996a), 'Husserl's Intersubjective Transformation of Transcendental Philosophy', *Journal of the British Society for Phenomenology*, 27/3, 228–45.
- Zahavi, Dan (1996b), *Husserl und die transzendente Intersubjektivität. Eine Antwort auf die sprachpragmatische Kritik* (Dordrecht: Kluwer).
- Zahavi, Dan (1999), *Self-Awareness and Alterity: A Phenomenological Investigation* (Evanston, IL: Northwestern University Press).
- Zahavi, Dan (2002a), 'Merleau-Ponty on Husserl: A Reappraisal', in Ted Toadvine and Lester Embree (eds), *Merleau-Ponty's Reading of Husserl* (Dordrecht: Kluwer), 3–29.
- Zahavi, Dan (2002b), 'Husserl und das Problem des vor-reflexiven Selbstbewusstseins', in Heinrich Hüni und Peter Trawny (eds), *Die erscheinende Welt. Festschrift für Klaus Held* (Berlin: Duncker & Humblot), 697–724.
- Zahavi, Dan (2003), *Husserl's Phenomenology* (Stanford, CA: Stanford University Press).
- Zahavi, Dan (2007), 'Killing the Strawman: Dennett and Phenomenology', *Phenomenology and the Cognitive Sciences*, 6/1–2, 21–43.
- Zahavi, Dan (2008), 'Internalism, Externalism, and Transcendental Idealism', *Synthese*, 160/3, 355–74.

11

Phenomenology and Transcendental Philosophy

Making Meaning Thematic

Steven Crowell

1. Introduction: Phenomenology and Phenomenologies

In the sense relevant to this chapter, ‘phenomenology’ is the philosophy of Edmund Husserl and of those philosophers who linked on to it by means of creative (even if often quite critical) appropriation. This defines a very large group, but among historical figures it includes at least Martin Heidegger, Hans-Georg Gadamer, Jean-Paul Sartre, Maurice Merleau-Ponty, Immanuel Levinas, and Jacques Derrida.¹ Husserl designated his mature thought ‘transcendental phenomenology’, but none of these philosophers adopted that designation for their version of phenomenology. Remarks can be found in the works of each that link them to the transcendental tradition, but in general the history of phenomenology appears to be a series of attempts to break free from the ‘intellectualism’ (Merleau-Ponty’s term) of transcendental philosophy.

Some might argue, then, that an examination of the relation between phenomenology and the transcendental turn inaugurated by Kant ought to restrict itself to laying out those aspects of Husserl’s thought that either draw upon or directly criticize tenets of Kant’s Critical philosophy, and several very good studies of this sort have been carried out.²

¹ The historical context is laid out in detail in Herbert Spiegelbert, *The Phenomenological Movement: A Historical Introduction*, third (revised and enlarged) edition (The Hague: Martinus Nijhoff, 1984). A more limited introduction, but one that highlights transcendental motifs, is Dermot Moran, *Introduction to Phenomenology* (New York: Routledge, 2000). For an attempt to specify what such thinkers have in common as phenomenologists, see Steven Crowell, ‘Is There a Phenomenological Research Program?’, *Synthese*, 131/3 (2002), 419–44.

² See, above all, Iso Kern, *Husserl und Kant: Eine Untersuchung über Husserls Verhältnis zu Kant und zum Neukantianismus* (The Hague: Martinus Nijhoff, 1964). Also instructive is Manfred Brelage, *Studien zur Transzendentalphilosophie* (Berlin: de Gruyter, 1965).

But even if later phenomenologists sought to distance themselves from Husserl, they often did so while adopting elements of his transcendental phenomenology. There are few studies that explore whether there might be aspects of transcendental phenomenology that are shared by these otherwise very different thinkers. The present chapter aims to contribute to this latter project.

A short essay cannot hope to be comprehensive in establishing such transcendental *bona fides*. The approach will thus be a strategic amalgam of the historical and the systematic. In historical perspective my focus shall be limited to Husserl and Heidegger against the background of Kant and neo-Kantianism. But this examination will be 'historical' only in a very attenuated sense, since I make no claim to exhaust even the most important historical connections relevant to an understanding of how Husserl and Heidegger belong to the transcendental tradition. Instead, the guiding thread will be systematic. I will argue that phenomenology—all phenomenology—is transcendental insofar as it *makes meaning thematic* as philosophy's primary *topos*. Taking as its theme not things but the meaning or intelligibility of things, phenomenology transforms transcendental philosophy by expanding its scope to embrace *all* experience, not just the cognitive, axiological, and evaluative 'validity spheres' addressed in Kant's three *Critiques*. Thus phenomenology accomplishes a universal generalization of the transcendental turn: inquiry into the (normative) conditions for the possibility of knowledge becomes an inquiry into *intentionality* or 'mental content' as such: our experience of something *as* something.

The phenomenological thematization of meaning involves (1) rejecting Kantian representationalism; (2) adopting the neo-Kantian idea that categories are normative; and (3) insisting on the first-personal character of philosophical method. The following section will suggest how these points emerge from Husserl and Heidegger's shared diagnosis of Kant's shortcomings. A similarly shared rejection of the neo-Kantian attempt to reformulate Kant's project as a theory of science provides the background for three subsequent sections which explore the phenomenological approach to meaning systematically from objective, subjective, and existential perspectives.

2. Phenomenology in Kantian Context

No doubt Dieter Henrich is right to insist on the ultimately practical or ethical motivation behind Kant's transcendental project:³ faced with his failure to bring the details of the Transcendental Deduction to complete clarity, Kant warned against taking such

³ I adopt the phrase 'transcendental project' from A. C. Genova's article, 'Good Transcendental Arguments', *Kant-Studien*, 75 (1984), 469–95, in which he argues that focus on transcendental arguments as the key to transcendental philosophy is misleading, since such arguments make sense only within Kant's overall critical *project*. This point is important for understanding how phenomenology, which unlike Anglo-American approaches to Kant pays little attention to transcendental arguments, can nevertheless be considered a continuation of Kant's project.

details for the heart of philosophy. That lay, rather, in justifying the ‘idea of freedom’ and providing a ‘defense of the essential interests of humankind.’⁴ Nevertheless, this defense demanded a theoretical alternative to metaphysical dogmatism and empiricist skepticism. Details aside, this alternative required a *Revolution der Denkungsart* that consists in positing a radical *discontinuity* between philosophy and all first-order cognition, whether empirical or metaphysical. Phenomenology is transcendental because it belongs to this revolution.⁵

The discontinuity thesis comes to expression in a passage in which Kant specifies what he means by ‘transcendental’. Asserting that ‘not every kind of knowledge *a priori* should be called transcendental, but that only by which we know that—and how—certain representations (intuitions or concepts) can be employed or are possible purely *a priori*,’⁶ Kant points out that transcendental philosophy is distinguished by the sort of question it asks. First-order inquiries—whether empirical like physics and psychology or *a priori* like mathematics and metaphysics—are carried out in an *intentio recta* and they establish the real properties of their objects. Transcendental critique, in contrast, asks how it is possible that such first-order thinking *can* yield knowledge, and it deals with objects and their properties only in a reflective *intentio obliqua* concerned with what makes them cognitively *accessible*. No first-order proposition can contribute to a transcendental account of how it is possible for any such first-order proposition to be an instance of cognition, nor can transcendental conditions of possibility be identified with some special realm of entities or some peculiar property of objects. As Kant put it: ‘The distinction between the transcendental and the empirical belongs . . . only to the critique of knowledge; it does not concern the relation of that knowledge to its objects.’⁷ To ask after transcendental conditions of knowledge is not to ask how particular judgments are formed; nor is it to establish whether any first-order judgment is true. These are tasks for first-order sciences. Instead, transcendental philosophy addresses how it is possible that the ‘ground’ appealed to in some first-order practice of justification—for instance, experience—can serve as a *norm* for such a practice.

Thus Kant’s project is not concerned with the real relation between a representation and its object but solely with the cognitive claim advanced in it, and the question of how knowledge is possible is not a factual but a normative one. It does not look for some causal connection between mind and world but investigates how a concept can *hold* of something—not ‘how can something represent an object?’ but ‘how can it

⁴ Dieter Henrich, ‘Identity and Objectivity: An Inquiry into Kant’s Transcendental Deduction’, in Dieter Henrich, *The Unity of Reason: Essays on Kant’s Philosophy*, ed. Richard Velkley (Cambridge, MA: Harvard University Press, 1994), 127.

⁵ Marks of this inclusion are phenomenology’s anti-naturalism and its metaphysical neutrality. On the former, Steven Crowell, *Husserl, Heidegger, and the Space of Meaning: Paths Toward Transcendental Phenomenology* (Evanston, IL: Northwestern University Press, 2001); on the latter, David Carr, *The Paradox of Subjectivity: The Self in the Transcendental Tradition* (Oxford: Oxford University Press, 1999).

⁶ Immanuel Kant, *Critique of Pure Reason*, tr. Norman Kemp Smith (London: Macmillan, 1968), 96 (A56/B80).

⁷ Kant, *Critique of Pure Reason*, 96 (A57/B81).

do so correctly?' Further, transcendental philosophy is concerned with the normative ground of knowledge precisely where that ground's claim to normative status involves a puzzle. For instance, the ground of synthetic judgments is experience, i.e. the direct intuition of an object. But because some synthetic judgments are known a priori, and because they are known by a subject whose intuition is exclusively sensible (receptive), the normative status of experience appears puzzling. How can it ground the universality and necessity characteristic of a priori knowledge?

So formulated, the transcendental question makes sense only against the background of some specific characterization of the cognitive situation of the knower—that is, a more or less explicit philosophical anthropology.⁸ This places a significant burden on any transcendental account of knowledge, for that account must be *consistent* with the anthropological picture. Rüdiger Bubner denotes this the 'self-referentiality' condition: 'that knowledge which is called transcendental takes as its object, together with the general conditions of knowledge, the conditions of its own genesis and functioning.'⁹ Any transcendental account of how we can know synthetic judgments a priori must also explain how such transcendental knowledge is itself possible. According to Husserl and Heidegger, however, Kant's picture of the knowing subject—his cognitive anthropology—renders a non-dogmatic account of the normative (categorical) conditions of cognition impossible.¹⁰

The problem surfaces at the heart of Kant's Transcendental Deduction of the categories. How are synthetic judgments possible a priori? What is the normative ground of the synthesis they assert? What accounts for the fact that they can be known to hold of objects? Kant answers with an argument that purports to show that such judgments are 'the apriori conditions of a possible experience in general' and are thus 'at the same time conditions of the possibility of objects of experience'.¹¹ Since we are interested here only in the self-referential aspect of Kant's position we need not linger over the details of his argument. The crucial question is this: what is it about the 'possible experience' mentioned in the first half of the cited passage that allows it to play a normative role in explaining how such judgments hold of objects?

⁸ This point informs John McDowell's strategic or 'therapeutic' transcendental approach, which begins with a contingent picture of our cognitive situation that makes it seem as though perception can play no justificatory role in knowledge and from there moves to replace that anthropological picture with another in which it is no longer puzzling. See John McDowell, *Mind and World* (Cambridge, MA: Harvard University Press, 1996).

⁹ Rüdiger Bubner, 'Kant, Transcendental Arguments, and the Problem of the Deduction', *Review of Metaphysics*, 28 (1975), 462.

¹⁰ Husserl speaks here of 'Kant's mythical constructions'—his '“faculties”, “functions”, “formations”', which we 'are unable to make intuitive to ourselves'. Edmund Husserl, *The Crisis of European Sciences and Transcendental Phenomenology*, tr. David Carr (Evanston, IL: Northwestern University Press, 1970), 114. Martin Heidegger, *Kant and the Problem of Metaphysics*, tr. Richard Taft (Bloomington, IN: Indiana University Press, 1997), 144ff., argues that 'the Anthropology worked out by Kant is an empirical one and not one which is adequate for the transcendental problematic, i.e., it is not pure'. He goes on to suggest that his own 'fundamental ontology' of Dasein is meant to address this problem.

¹¹ Kant, *Critique of Pure Reason*, 138 (A111).

Here Kant's anthropological presuppositions prove decisive. Descriptively, Kant has previously established that experience involves two independent 'sources', sensibility and understanding. But by themselves these sources do not suffice, for experience also involves the (descriptive?) fact of being *owned*, mine. This owned character of experience—the 'transcendental unity of apperception', the 'abiding and unchanging "I"', or 'self-consciousness'—is the '*absolutely* first and synthetic principle of our thought', and it accounts for 'the synthetic unity of the manifold in all possible *intuition*'.¹² Thus, anything that can be shown to be necessary for this apperceptive consciousness of unity is thereby also shown to be necessary for any object that can be 'represented' in it.

The problem with this is that it does not satisfy the self-referentiality condition: the concept of 'possible experience' which the Deduction establishes as the scope of synthetic a priori knowledge is so restrictive that transcendental critique cannot count as knowledge. Knowledge of the necessarily 'owned' character of experience is neither a matter of logic nor 'an inner experience'; it is an 'apperception' that 'does not know itself through the categories'.¹³ If one were justified in assuming an apperceptive grasp of *necessary* ego-unity, one could treat it as a principle by which to delimit conditions of possible experience, but Kant provides no non-dogmatic way to justify such an assumption. The ultimate principle of the deduction remains 'a *thought*, not an *intuition*',¹⁴ and its claim to normative status as the ground of philosophical knowledge remains a puzzle.

The effect of this objection is to force reconsideration of the anthropological horizon of Kant's theory, and here two very different paths open up. One attempts to preserve the role of subjectivity in transcendental philosophy by improving upon Kant's understanding of it, while the other rejects any such role. The first leads to phenomenology and to some naturalistic, cognitive-science readings of Kant,¹⁵ while the second leads from Hegel to neo-Kantianism and to an interest in transcendental arguments. The first takes its departure from the A-Edition version of Kant's transcendental deduction, with its account of how categories originate in a threefold synthesis (apprehension, reproduction, recognition). This may be called the 'psychological' reading, since it attributes syntheses other than purely inferential or logical ones to the transcendental subject. The second—which may be called the 'logical' reading—emphasizes the B-Edition's insistence on the autonomy of the understanding (i.e. the purely inferential, categorial character of the experiential synthesis) and the merely 'formal' character of the unity of apperception.

Through Hermann Cohen's work, the logical reading dominated the Marburg School of neo-Kantianism, which rejected the genetic 'subjective' deduction of the categories

¹² Kant, *Critique of Pure Reason*, 142 (A117–18); 146 (A123).

¹³ Kant, *Critique of Pure Reason*, 329 (A342/B400); 365 (A402).

¹⁴ Kant, *Critique of Pure Reason*, 168 (B157).

¹⁵ See, for instance, Patricia Kitcher, *Kant's Transcendental Psychology* (Oxford: Oxford University Press, 1990).

as well as the supposedly psychologistic assumption that sensible intuition provides a contribution to knowledge independent of the categories.¹⁶ It was against such proposals that Husserl developed his phenomenological version of transcendental philosophy, and against them, too, that Heidegger directed his own Husserl-inspired Kant-interpretation.¹⁷ In doing so they adopted the psychological reading, broadly conceived; that is, they sought to clarify the normative conditions of experience by reflecting on 'consciousness' or 'being-in-the-world'. Yet phenomenology is no less committed than neo-Kantianism to the view that the transcendental subject is not the psychological subject embedded in the causal nexus of empirical reality. The difficulty, then, is to understand how the transcendental subject can be concrete—more than the neo-Kantian formal principle of unity—yet not a nexus of psychological causality and association.

Klaus Hartmann, a contemporary neo-Kantian, argues that this difficulty indicates a fatal flaw in what he calls 'mixed' transcendental theories, of which both Kant and Husserl offer versions. Mixed theories combine—and ultimately confuse—elements of a reflection on the *constitution* of knowledge with elements of a transcendental *justification* of knowledge.¹⁸ Constitutional reflection starts with 'realist' affirmations concerning 'our' nature which, as regards their own justification, are mere matters of fact, genetic assumptions, or 'metaphorical ordinary-level conceptions'.¹⁹ From this constitutional level, mixed theories try to construct a bridge to a priori or categorial necessity. Kant, for instance, distinguishes between 'pure' and 'empirical' intuition, 'pure' and 'empirical' synthesis, etc. But such strategies always fail: the genetic assumptions are simply disguised as 'a priori conditions of knowledge', though we in fact 'do not understand how acts are grounds for truth'. The bottom line is that 'transcendental acts and syntheses . . . cannot account for the validity of knowledge'.²⁰

Phenomenology answers this charge with another: 'pure' category theories must be *dogmatic*. Hartmann himself admits that a theory, like Hegel's, which 'makes [no]

¹⁶ On the Hegelian background of Marburg neo-Kantianism, see Klaus Christian Köhnke, *The Rise of Neo-Kantianism: German Academic Philosophy between Idealism and Positivism*, tr. R. J. Hollingdale (Cambridge: Cambridge University Press, 1991). On its formalism, see Brelage, *Studien zur Transzendentalphilosophie*. Contemporary philosophers who can be located in this Hegelian lineage include Robert Pippin, *Hegel's Idealism: The Satisfactions of Self-Consciousness* (Cambridge: Cambridge University Press, 1989); Robert Brandom, *Making it Explicit* (Cambridge, MA: Harvard University Press, 1994); McDowell, *Mind and World*; and Klaus Hartmann, *Studies in Foundational Philosophy* (Amsterdam: Editions Rodopi, 1988).

¹⁷ Martin Heidegger, *Phenomenological Interpretation of Kant's Critique of Pure Reason*, tr. Parvis Emad and Kenneth Maly (Bloomington, IN: Indiana University Press, 1997), 292: 'When some years ago I studied the *Critique of Pure Reason* anew and read it, as it were, against the background of Husserl's phenomenology, it opened my eyes; and Kant became for me a crucial confirmation of the accuracy of the path which I took in my search.'

¹⁸ Hartmann, 'On Taking the Transcendental Turn', in *Studies in Foundational Philosophy*, 229; also Hartman, 'Transcendental Argumentation—Options and Preferences', in *Studies in Foundational Philosophy*, 244.

¹⁹ Hartmann, 'On Taking the Transcendental Turn', 204.

²⁰ Hartmann, 'On Taking the Transcendental Turn', 200; 'Transcendental Argumentation', 249.

concessions to existential considerations' suffers from 'circularity'.²¹ Instead of ignoring the anthropological or subjective, then, phenomenology reconceives its function: it does not, as it does for Kant, serve as a *premise* in a transcendental argument intended to establish the validity of the categories; it serves, rather, as philosophy's primary datum—intentionality and its categorial meaning-structures reflectively grasped. For phenomenology, the puzzling relation between the normative and the factual is found in all meaningful experience, all experience of something as something. Thus it transforms transcendental philosophy from an epistemic project of justifying certain a priori principles into a (so to speak) 'semantic' project of clarifying intentionality through reflection on the evidence of first-person experience.²²

In unpacking some implications of this transformation I shall first examine the phenomenological approach to the intentional *object*, in particular its critique of representationalism. Second, I shall take up the *noetic* aspect of intentionality, contrasting the phenomenological approach with the categorial *Geltungstheorie* of the neo-Kantian philosopher, Emil Lask. Finally, I shall explore Heidegger's claim that the transcendental account of intentionality requires an existential turn.

3. The Intentional Object and the Critique of Representationalism

To open transcendental philosophy to the full range of intentional experience is, in Husserl's terms, to investigate the conditions that make 'transcendence'—consciousness of the real in the widest sense—possible.²³ Heidegger introduces a distinction here that will prove important later on, between 'ontic transcendence' and 'original' or 'primal' transcendence. Ontic transcendence is equivalent to intentionality: 'Intentionality is indeed related to the beings themselves and, in this sense, is an ontic transcending comportment'. In it the subject 'transcends' itself toward beings other than itself. But, Heidegger continues, 'the problem of transcendence as such is not at all identical with the problem of intentionality. As ontic transcendence, the latter is itself only possible on the basis of original transcendence, on the basis of

²¹ Hartmann, 'Hegel: A Non-Metaphysical View', in *Studies in Foundational Philosophy*, 282, 280. See also 'On Taking the Transcendental Turn', 217: 'Accordingly, falsifiability in transcendental philosophy poses a special problem.'

²² For extensive discussion of the implications of this shift see J. N. Mohanty, *The Possibility of Transcendental Philosophy* (The Hague: Martinus Nijhoff, 1985).

²³ Edmund Husserl, *The Idea of Phenomenology*, tr. William P. Alston and George Nakhnikian (The Hague: Martinus Nijhoff, 1964), 27: 'If we look closer at what is so enigmatic and what, in the course of subsequent reflection on the possibility of cognition, causes embarrassment, we will find it to be the transcendence of cognition. All cognition of the natural sort, and *especially the pre-scientific* [my emphasis], is cognition which makes its object transcendent. It posits objects as existent, claims to reach matters of fact that are not "strictly given to it", are not "immanent" to it.' For Husserl, transcendental knowledge is any knowledge that contributes to clarifying the 'enigma' of transcendence.

being-in-the-world. This primal transcendence makes possible every intentional relation to beings.²⁴ To see why this is so, it is first necessary to see what a transcendental clarification of intentionality involves.

Intentionality, ontic transcendence, is the experience of something as something. Thus all intentional states involve a meaning (the 'as') through which some object is purportedly present to consciousness. But in everyday life consciousness is directed toward the object, not the meaning. To make this meaning thematic, Husserl focuses first on the 'natural attitude': in the straightforward gearing into the world characteristic of everyday life I busy myself with things that come to my attention, taking them as simply 'there' just as they present themselves. This natural attitude thus involves a pervasive belief in the reality of what shows itself, a belief whose correlate is not this or that aspect or thing but the *world*, the horizon of all such particular experiences: "The world is always there as an actuality; here and there it is at most 'otherwise' than I supposed; this or that is, so to speak, to be struck *out of it* and given such titles as 'illusion', and 'hallucination', and the like . . ."²⁵ Such realism is what makes the natural attitude a topic for transcendental inquiry.

To speak of 'realism' here is to acknowledge that my experience of the world involves a claim to validity and so a normative distinction between success and failure. What presents itself is not only there as *something* but also as 'truly' or 'validly' existing. Transcendental inquiry is concerned with how such a normative claim—this consciousness of *true* being—is grounded. Indeed, experience in the natural attitude is characterized by two distinct normative moments: the (defeasible, hence norm-governed) claim to experience what *truly* is, and also the (defeasible, hence norm-governed) claim to experience something of just this *sort*. The act of perception in which I experience something as a tree, for instance, somehow involves reference to normative satisfaction conditions that determine what the thing is *supposed* to be, a certain meaning that establishes how the current experience must necessarily be related to (former and) subsequent experiences *if* the thing perceived is in fact what it gives itself to be—namely, a tree.²⁶ More generally, we may speak of the 'intentional content' of experience only where it makes sense to talk about how such experience might go wrong. The puzzle—that which demands transcendental clarification—is how this normatively structured content of experience is possible. What *is* 'content', or meaning, anyway?

²⁴ Martin Heidegger, *Metaphysical Foundations of Logic*, tr. Michael Heim (Bloomington, IN: Indiana University Press, 1984), 135.

²⁵ Edmund Husserl, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, First Book*, tr. F. Kersten (The Hague: Martinus Nijhoff, 1983), 57.

²⁶ On conditions of satisfaction see John Searle, *Intentionality* (Cambridge: Cambridge University Press, 1983), 10–11. Searle's conditions are logical not phenomenological, i.e. they *reconstruct* experience without having to be evident *in* experience. But as Hubert Dreyfus argues, rational reconstruction presupposes that one first get the phenomenology right. See Dreyfus, 'Reply to John Searle', in Heidegger, *Coping, and Cognitive Science: Essays in Honor of Hubert L. Dreyfus*, ii, ed. Mark Wrathall and Jeff Malpas (Cambridge, MA: MIT Press, 2000), 323–37.

In everyday life—and so also in all first-order inquiries—consciousness is directed toward objects and their real properties; that is, toward leafy trees, heavy hammers, and so on. The meaning thanks to which such objects are there for us as trees and hammers, however, is no property of the tree or the hammer. Nor can it be an object immanent to consciousness, a ‘representation’ grasped in psychological reflection, for any such immanent object would have to be given through a meaning, and so on *ad infinitum*. Thus, while it might appear that Kant’s concept of *Vorstellung* (representation)—which includes both concepts and percepts (intuitions)—suffices to capture the intentionality of experience, the notion of representation already conceals the phenomenon of intentionality by treating it as a relation between two entities, one subjective and the other objective. The phenomenological approach to meaning avoids the pitfalls of a theory of representation in the modern sense of the ‘way of ideas’.²⁷ In contrast to all such theories, phenomenology does not locate meaning in the subject but identifies it ‘transcendentally’ with the object. As Husserl writes:

In a certain way, and with some caution in the use of words, we can also say that *all real unities are ‘unities of meaning’* . . . Reality and world are names precisely for certain valid unities of meaning . . . related to certain concatenations of absolute, of pure consciousness which, by virtue of their essence, bestow meaning and demonstrate meaning-validity precisely thus and not otherwise.²⁸

The intentional object as it is considered in phenomenological reflection—that is, in a reflection that makes explicit the meaning through which the object is given—is called the ‘noema’. The noema has been interpreted in Fregean terms as an abstract entity that mediates between consciousness and the external world, but this wreaks havoc with Husserl’s anti-representationalism. It is better seen as the object itself considered in terms of its modes of self-givenness.²⁹ Noematic unities are not unities of objective properties but of ways in which objects and their properties are given. Nor are these ‘ways’ or modes of givenness mental items that get grouped causally or associatively. From such association no valid consciousness of transcendent objects could ever arise. To yield object-consciousness, modes of givenness must be related to one another through something like a *rule*: one mode of givenness ‘implies’ another, and it is the consciousness of such ‘intentional implications’ that allows my experience to be an experience of *this* object, an objective identity of a certain sort.

²⁷ Of course, there are other uses of the term ‘representation’ that do not involve this problem—for instance, it is sometimes used merely to indicate *that* things are present to me in a certain way or under a certain description. But Kant’s usage seems not to be of this sort, and we shall restrict ourselves here to representation taken as a mental entity. Whether this is a fair reading of Kant may be questioned, but we cannot enter into the issues.

²⁸ Husserl, *Ideas I*, 129.

²⁹ A lively debate around this issue was inaugurated by Dagfinn Føllesdal’s ‘Husserl’s Notion of Noema,’ *Journal of Philosophy*, 66 (1969), 680–7. The essays collected in Hubert L. Dreyfus and Harrison Hall (eds), *Husserl, Intentionality, and Cognitive Science* (Cambridge, MA: MIT Press, 1982) generally accept Føllesdal’s interpretation and explore its ramifications, while a more critical account of that interpretation, together with a positive alternative which I generally—if not in every detail—follow here is provided by John Drummond, *Husserlian Intentionality and Non-Foundational Realism: Noema and Object* (Dordrecht: Kluwer, 1990).

For instance, reflective description of my perception of the file cabinet in my study does not reveal a series of 'sense data' that are 'formed' by a rule imposed by the mind. The rule that establishes its reality is inherent in the perception itself. To see a *cabinet* from here is to see its visible front side as entailing a back side that is not now seen but would become visible in certain quite specific ways were I to change my location. To say that perception of the front 'entails' a back side means that the link between modes of givenness is norm-governed: were I to move to get a 'better' look at the back and discover that there was none (as in a hologram, for example), I would experience the 'collapse' of my object, that is, the invalidity of its claim to be a cabinet. Were experience not governed in this normative way, subsequent experience could never lead me to revise my previous experience; it would simply replace it. The intentional object, then—the noema, the object as it is experienced—is a normatively structured unity of meaning.

Yet the claim that things in the world are unities of meaning is hardly an obvious one. If one understands it as a first-order claim, it will seem to entail either a subjective or a metaphysical idealism. It is, however, a *transcendental* claim; that is, it 'belongs only to the critique of knowledge' and 'does not concern the relation of that knowledge to its objects.'³⁰ Phenomenology establishes this point—the transcendental discontinuity thesis—by means of the phenomenological reduction or *epoché*, the significance of which can be appreciated by returning to Husserl's description of the natural attitude.

Everyday life is characterized by a kind of global realism, a belief in the factual existence of what I encounter. Husserl calls this the 'general positing' of the natural attitude. As noted above, this involves a validity claim distinct from the norms that determine one's experience as being an experience of this *sort* of thing. Thus I can focus on the thing's ways of being given while 'bracketing' the existential validity claim. For instance, I need take no stand on whether my belief in the reality of the file cabinet I perceive will be borne out; that is, I can 'suspend' my 'positing' of the cabinet as existing, take no stand with regard to its claim to be. And if I can do so in the case of this or that entity, I can do so universally: 'We put out of action the general positing which belongs to the essence of the natural attitude; we parenthesize everything which that positing encompasses with respect to being.' Thereby, Husserl continues, 'I am not negating this "world" as though I were a sophist; I am not doubting its factual being as though I were a skeptic; rather I am exercising the "phenomenological" *epoché* which also *completely shuts me off from any judgment about spatiotemporal factual being*.'³¹ The *epoché* does not deny the being of the world; rather, it is a way of allowing the meaning of that being, its as-structure, to become thematic.

With respect to the phenomenological critique of representationalism, the most important point about the *epoché* is that it sharply distinguishes phenomenology from all naturalistic (or metaphysical) explanations of meaning. Because I take no stand

³⁰ Kant, *Critique of Pure Reason*, 96 (A57/B81).

³¹ Husserl, *Ideas I*, 61.

with regard to factual being, I cannot make any judgments that presuppose such being. Hence I must 'exclude all sciences relating to this natural world . . . I make absolutely no use of the things posited in them. Nor do I make my own a single one of the propositions belonging to those sciences'.³² For this reason, phenomenology can make no use of the concept of representation, that is, a mental entity that is posited by psychology to explain why things show up in the world as meaningful. Descriptively, there is no mental entity that mediates between the act (mind) and its object (world), and since the *epoché* precludes causal explanation, phenomenology has no reason to posit one. Nor is there any reason to posit a *Ding an sich*. The intentional object is not something subjective; intentional content is not a function of how we represent the world but of how the object *presents* itself. To characterize objects transcendently as unities of meaning, then, is to connect with what the earlier metaphysical tradition called 'being in the sense of truth' (*on hos alethes; ens tanquam verum*): meaning is the object, what the thing is 'in truth'.

The phenomenological reduction thus involves a rejection of what John McDowell calls the 'sideways on' view of our experience.³³ There are various ways in which an object is there for me in experience, but these ways are not mental representations of the sort about which one might ask whether they could *ever* be veridical. The only justificatory questions that the *epoché* leaves open are those first-order questions that arise within ordinary experience: Are the intentional implications that normatively structure the experience of some object fulfilled or disconfirmed by further experiences of the sort demanded by that kind of object? Is my perception of a snake in the corner confirmed by what I see or hear when I move closer, or did I not see a snake, but a rope? These, obviously, are not questions for philosophy to decide; the philosophical task—one that the phenomenological reduction makes possible—is to elucidate the normative structure of meaning, or intentional content, that is presupposed when such justificatory questions are negotiated in experience itself.

Here, neo-Kantians like Klaus Hartmann and Heinrich Rickert object that phenomenology, so described, has abandoned the very question that transcendental philosophy was supposed to answer.³⁴ If phenomenology does nothing more than describe the interplay among intentional contents within first-person experience, then it is in no position to answer the *questio juris*, that is, to explain how we move from reflectively describing truth-claims to establishing that and how such claims can deliver *truth*.

It should be admitted that such critics have a point. Phenomenology provides no principle for deducing that some particular experience (or kind of experience) is veridical. Phenomenology does not aim to answer the skeptic, however, but to undercut the motives for skepticism through 'presuppositionless' description of the

³² Husserl, *Ideas I*, 61.

³³ McDowell, *Mind and World*, 34–6.

³⁴ Heinrich Rickert, *Die Philosophie des Lebens: Darstellung und Kritik der philosophischen Modeströmungen unserer Zeit* (Tübingen: J. C. B. Mohr, 1922). Also Rickert, 'Zwei Wege der Erkenntnistheorie: Transscendentalpsychologie und Transscendentallogik', *Kant-Studien*, 14 (1909), 189–93.

entailments that make up the *logos*, so to speak, of experience itself.³⁵ For instance, only if one's description of the normative structure of experiential confirmation and disconfirmation is burdened with metaphysical or naturalistic assumptions about subject and object will skeptical arguments based on the idea of the mind as a *forum internum* or on global extension of the argument from illusion seem compelling. In phenomenological perspective, failure to recognize this point vitiates the neo-Kantian attempt to answer the *questio juris* with a theory of categories that suppresses any 'psychological' reflections on subjectivity.³⁶ To see why reflection on subjectivity is necessary, we turn to the noetic side of intentionality and examine Husserl's claim that unities of meaning require a 'pure, meaning-bestowing consciousness'.

4. Subjectivity: Phenomenology between Logic and Psychologism

We may approach this issue by taking a closer look at a 'pure' category theory that, like phenomenology, breaks with Kant's representationalism and thematizes meaning as a transcendental concept, but, unlike phenomenology, sees no need for reflection on intentional acts. Emil Lask, a student of Heinrich Rickert, arrives at Husserl's concept of meaning as being in the sense of truth, but by a different route. Lask does not start with intentionality, the way objects are given in experience, but with a problem internal to Kant's transcendental logic, namely, the ontological status of categories. On the psychological reading, which Lask rejects, categories are understood as real mental processes, 'forms' that subjectively 'synthesize' the data of intuition. Such a view cannot explain how categories have cognitive import. But what status do categories have on the logical reading? 'The logical is precisely the logical and neither metaphysical nor psychological. But what sort of thing is it then?'³⁷ According to Lask, Kant never answers this question. Logical categories remain 'homeless' in his version of the 'two-world' system, since they are neither functions of the mind in the psychological sense nor of things in themselves in the metaphysical sense.³⁸

³⁵ On phenomenology's stance toward skepticism, see J. N. Mohanty, 'Rorty, Phenomenology, and Transcendental Philosophy', in *Possibility of Transcendental Philosophy*, 57–66.

³⁶ Against the neo-Kantian emphasis on the *questio juris* Heidegger, *Phenomenological Interpretation of Kant's Critique of Pure Reason*, 223–4, writes: 'If we were to remain within the Kantian terminology, then we would have to say that precisely *not* a *questio juris* but a *questio facti* lies at the center of the problem of the transcendental deduction'—namely, the question of the transcendental constitution of the subject. Heidegger continues: 'Kant speaks of two sides of the transcendental deduction, a subjective one and an ob-jective one.' But 'he fails to see that by radically carrying out the subjective side of the task of deduction, the ob-jective task is taken care of'.

³⁷ Emil Lask, *Die Logik der Philosophie und die Kategorienlehre*, in *Sämtliche Werke, Zweiter Band* (Jena: Dietrich Scheglmann Reprintverlag, 2003), 24. This edition provides, in square brackets in the text, the pagination of the standard edition of Lask's work: Emil Lask, *Gesammelte Schriften*, ii, ed. Eugen Herrigel (Tübingen: J. C. B. Mohr, 1923). Henceforth references will be given with the page number of the *Sämtliche Werke* edn first, followed by that of the *Gesammelte Schriften* edn, e.g. 24/26. All translations from Lask are my own.

³⁸ Lask, *Logik der Philosophie*, 12/14, 110/131.

For this reason Kant fails the self-referentiality test: 'In Kant's theory of categories there is no place for the categorial forms of his own speculation, and thus the critic of theoretical reason denies the logical conditions of his own critique of reason.'³⁹

To address this lacuna, Lask highlights the normative character of categories, which provides the basis for a genuinely transcendental two-world theory. In transcendental logic, the 'totality of the thinkable' is not divided into two realms of entities—mind and world—but into the realm of entities as such (physical, psychical, metaphysical, etc.) and the world of logical forms that constitute their 'clarity' or 'intelligibility'.⁴⁰ Following Hermann Lotze, Lask distinguishes between 'what is and what is valid, the domain of being and the domain of validity, ontic constructs and valid constructs, between the sphere of reality and the sphere of value, between that which *is* there and *occurs*, and that which *holds* without having to be.'⁴¹

To say that categories are 'validities' (*Geltendes*) is to say that they hold. To hold of something is to be true of it; categorial form is the 'truth content' (*Wahrheitsgehalt*) of the object. Thus Lask arrives at a transcendental conception of the object similar to Husserl's: an object is constituted by categorial form and that of which it holds; that is, it is a unity of meaning: '*Meaning (Sinn)* shall denote the unity, the combination of form and material, the whole which consists of the in itself empty and dependent form together with its fulfilling content. The realm of objects . . . is a realm of "meaning"'.⁴²

On such a view, the entities that in pre-Kantian philosophy make up the totality of what is are merely the *material* for objects in the transcendental sense. For instance, 'being the cause of the water's boiling' is not a property (even a relational one) of the flame on my stove. Rather, both the boiling water and the flame are the material of which the category 'causality' holds. To hold is to render intelligible, to 'illuminate' how it stands with things. As validities, categories have no independent existence; they are *nothing but* certain ways of being of the material, certain 'involvements' (*Bewandtnisse*): terms for logical categories such as 'objectivity, being, . . . reality, existence', refer to the 'particular objective involvement' characteristic of some specific range of material.⁴³ By means of such objective involvements the material already stands in the normative space of reasons, categorial space, before I make any judgment about it.

Lask's definition of the object as meaning makes no reference to subjective syntheses or experiences. Nevertheless, the object is not a thing in itself, something that transcends all possible experience. It is *erlebt* before it is *erkannt*; experience is an 'immediate living in truth'.⁴⁴ In everyday life we are focused on the object material while the object in the strict sense, meaning, is merely lived *through*. In both everyday and scientific inquiry we make the involvements of the material explicit—for instance, we discover causal connections or determine the real properties of things—but it is only in

³⁹ Lask, *Logik der Philosophie*, 216/263.

⁴⁰ Lask, *Logik der Philosophie*, 64/75.

⁴¹ Lask, *Logik der Philosophie*, 5/5.

⁴² Lask, *Logik der Philosophie*, 30/34.

⁴³ Lask, *Logik der Philosophie*, 59/69.

⁴⁴ Lask, *Logik der Philosophie*, 160/192.

transcendental philosophy that the object is grasped explicitly as meaning. Thus both Lask and Husserl expand the scope of transcendental philosophy to include the satisfaction conditions, or categories, that constitute the intelligibility of pre-theoretical experience: 'the most basic logical problems disclose themselves only to a researcher who also incorporates "pre-scientific" cognition into the domain of his investigation'.⁴⁵ But this opens Lask's supposedly pure category-theory (transcendental *logic*) to the problem of intentionality, thereby revealing its implicit dogmatism. For there simply *is* no approach to "pre-scientific" cognition—to the normative structure of what passes below the radar of explicit scientific judgments—without a reflective description of the various experiences themselves. For Lask, the 'genuinely transcendent' object 'untouched by all subjectivity' is the sole concern of transcendental logic. But without reflection on intentionality one can only *posit* some set of categories that one imagines, for whatever reason, to hold of pre-scientific cognition. To talk of objects as constituted by logical form is empty without an account of how their truth content, their objective involvement, shows itself in experience. Transcendental logic must finally be grounded in transcendental phenomenology.

Husserl and Heidegger offer detailed and complementary accounts of such pre-theoretical involvements. As we saw, the object can be understood phenomenologically as an open set of intentional implications, anticipations of further experiences governed by norms of confirmation or disconfirmation (truth). But these implications are discernable only if one specifies the acts (in Husserl's terms, the 'noeses') of which they are the correlates. For instance, the front side of my cabinet does not of itself entail a back side: there *is* no front or back of the cabinet without reference to some act of perception. It is this act that determines the precise sort of confirmation or disconfirmation at issue in the constitution of an object.⁴⁶ A perception is confirmed or disconfirmed by the further perceptions of 'the same' thing that are entailed in it, but a memory is confirmed by the evidence of the past perception it entails, not by another memory. For philosophers like Lask and Klaus Hartmann, reflection on acts produces a mixed transcendental theory which is compromised by empirical or psychological elements. But phenomenological act-analysis is not psychological; rather, it thematizes the *norm-governed* (Husserl calls them the 'teleological') relations that hold among acts: phenomena are 'ordered in an overall connection, in a "monadic" unity of consciousness, a unity that in itself has nothing at all to do with nature, with space and time or substantiality and causality, but has its thoroughly peculiar "forms"'.⁴⁷ It is only from such a perspective that the involvements of the noema come explicitly to light.

⁴⁵ Lask, *Logik der Philosophie*, 154/186.

⁴⁶ Mark Sacks has emphasized just this point in his account of how transcendental arguments function by moving from the conceptual claim involved in a judgment to the 'situated thought' that is presupposed in making it and that provides 'the apriori ground for its being true'. See Mark Sacks, 'The Nature of Transcendental Arguments', *International Journal of Philosophical Studies*, 13/4 (2005), 434–60, esp. 443–4. See also Mark Sacks, *Objectivity and Insight* (Oxford: Oxford University Press, 2000).

⁴⁷ Edmund Husserl, 'Philosophy as Rigorous Science', in *Phenomenology and the Crisis of Philosophy*, tr. Quentin Lauer (New York: Harper & Row, 1965), 108.

One of the categorial involvements of my cabinet, for instance, is its being a thing-with-properties, of which I show myself aware when I move around it to get a better look. To be aware, in this sense, is neither merely to register data as film registers light; nor is it to make a judgment about a state of affairs. Rather, on Husserl's account, it consists of a synthesis of noetic moments that are 'founded' (*fundiert*) on one another—that is, exhibit asymmetrical entailment relations. Perception presents the cabinet as there all at once (not given in phases that succeed one another, as notes in a melody) and as having certain properties. Nevertheless, this 'simple' givenness involves a norm-responsive synthesis of perceptual acts upon which it is founded. To see the cabinet as black, for instance, means that my perception is beholden to a norm of 'appearing better', that it anticipates a view in which the 'true' black presents itself and of which its actual appearance now—not so much black, but 'as black would look under the currently bad lighting in my office'—is a proper adumbration. The blackness of the cabinet *must* look silverish in this current act of perception, and it *must* change in a very specific way if I move to a different part of the room, or if the lighting changes. To perceive the noematic involvement captured by the category 'objective property' is possible in no other way. Without referring to the rule-governed relation between noeses (the various acts of perception in which the 'same' black color is given), the cabinet's color would appear to change with each new view, and nothing like an objective property could be experienced.⁴⁸

Relations of noetic founding are also constitutive of pre-theoretical involvements such as the cabinet's utility, of which I show myself aware when I pick up the files in order to clear my desk. Seeing the messy desk founds an affect of disgust (i.e. I am disgusted *with* the messy desk), and these together found an act of desire—for a clean desk, for a place where the files will be safe from spilled coffee. On this basis that thing next to the desk, the cabinet, is disclosed as potentially satisfying those desires.⁴⁹ Such desire may in turn found an act of willing, thanks to which an item on the front of the cabinet is disclosed as a handle to be seized and pulled. These acts are not carried out successively. None is an explicit judgment, and the relations between them are not logical (there is no purely logical connection between, say, perceiving and desiring). Yet their synthesis in my awareness of the cabinet's utility is not a matter of simple association or causality. They are bound by a phenomenologically irreducible kind of entailment. 'Pure' or transcendental consciousness—consciousness as thematized under the reduction—can thus be called 'meaning-bestowing' not because it imposes form on formless material, or because the properties of things are ontologically relative to my

⁴⁸ See John Drummond, 'On Seeing a Material Thing in Space: The Role of Kinaesthesia in Visual Perception', *Philosophy and Phenomenological Research*, 40 (1979–80), 19–32.

⁴⁹ In Husserlian terms, to say that an act of desiring is founded on a perceptual act is not to say that I first have one and then the other. Rather, it is to say that any instance of desire makes reference to (or includes within its intentional essence) something desirable, and what this something is refers to an act of perception. On this view, perception is not limited to sense perception but includes any act in which something is given 'in person', is *there* as itself. Thus numbers, no less than trees, are perceived, though not sensuously.

acts, but because specific involvements of the noema can show themselves only on the basis of founded relations between noeses.

If one looks more closely at the nature of these relations, however, it becomes questionable whether transcendental subjectivity should be identified with consciousness. Certainly, consciousness is a necessary condition for intentionality. But talk of 'acts' and 'syntheses' becomes strained when called upon to account for intentional content in the widest sense. In his critique of Husserl, Heidegger argues that acts of pure consciousness are abstractions; their 'thoroughly peculiar "forms"', or normative structure, cannot be understood apart from the practical contexts in which the subject, as embodied agent, is engaged.⁵⁰ Thus in his own reflection on the involvements that constitute the object as what it is, Heidegger avoids the term 'consciousness' and defines the transcendental subject as 'care' (*Sorge*).⁵¹ This is no mere terminological difference.

For Heidegger, categorial involvements cannot be understood in isolation but only as part of a 'totality of involvements' (*Bewandtnisganzheit*) that takes the form of a complex of 'in-order-to' relations.⁵² The noetic correlate of this totality is not an individual act such as perception or desire but a kind of practical comportment. My cabinet presents itself as in-order-to hold files;⁵³ that is, it is encountered in light of what is appropriate, of what a cabinet is *supposed* to be, a condition that it can succeed or fail in satisfying. But this sort of condition can belong to no act of perception, desire, or will—nor to any combination of them—if by 'act' one means, as Husserl did, something like a mental process or propositional attitude. A norm of what is appropriate can be present in consciousness only because it is first there in the exercise of certain abilities, skills, and instituted practices. Grasping the handle in order to open the cabinet involves an act of will, but it is my ability to grasp the handle that distinguishes such willing from mere wishing. The normative moment that constitutes the cabinet as useful for clearing my desk does not arise from my desire to clear my desk; rather, the meaning of that desire points back to the proprieties belonging to the skill I exercise in writing, consulting notes, etc., in light of which the crowd of things

⁵⁰ Thus, after a lengthy and quite appreciative interpretation of Husserl's analysis of intentional acts, Heidegger insists that this still does not get at the 'entity which is intentional'. See Martin Heidegger, *History of the Concept of Time: Prolegomena*, tr. Theodore Kisiel (Bloomington, IN: Indiana University Press, 1985), 110. Husserl understood this point as well: transcendental subjectivity is embodied, intersubjective, temporal/historical, and practical. See Dan Zahavi, *Husserl's Phenomenology* (Stanford, CA: Stanford University Press, 2003). However, Husserl never fully reconciled his concept of 'absolute' consciousness with this more full-blooded notion of subjectivity.

⁵¹ For an excellent discussion of this move from the perspective of the transcendental tradition, see Carl-Friedrich Gethmann, *Verstehen und Auslegung: Das Methodenproblem in der Philosophie Martin Heideggers* (Bonn: Bouvier Verlag, 1974).

⁵² Heidegger, *Being and Time*, tr. John Macquarrie and Edward Robinson (New York: Harper & Row, 1962), 114–17 (H83–5).

⁵³ It is important to note that for Heidegger this self-presentation is not a matter of becoming explicit or thematic. The genuine self-showing of things precedes 'attending' to them: 'The peculiarity of what is proximally ready-to-hand is that, in its readiness-to-hand, it must, as it were, withdraw in order to be ready-to-hand quite authentically' (Heidegger, *Being and Time*, 99 (H69)).

on the desk is a 'hindrance'. These skills, in turn, are determined *as* skills within specific practices—such as office work and data preservation—situated within institutions governing the symbolic and legal status of certain scraps of paper, deadlines, and so on. Even perceptual content cannot be clarified solely with reference to acts of perception. When I perceive the cabinet, an unseen back side is entailed only because there is a conditional reference to my ability to walk around the thing, or to keep my eye on it while it moves. Such content also refers to communicative practices, since I could not experience the unseen back side as there *now* without reference to what others could *report* seeing from behind the cabinet while I remained in front. Heidegger's phenomenology thus suggests how 'practical intentionality' provides the conditions of possibility for 'act intentionality', for meaning conceived as mental content.

Yet there must be more to the story. If practical intentionality—the skillful ability to accomplish certain tasks or to act in accord with the norms of a practice—were sufficient for meaning, a robot would be capable of encountering something *as* something.⁵⁴ According to Heidegger, the possession of intentional content requires a further condition: I must be able to understand *myself as being* up to something at which I might succeed or fail. Thus Heidegger, no less than Husserl, acknowledges the phenomenological importance of subjectivity, the first-person perspective, as a transcendental condition on meaning. But we would move in a circle were we to hold that such self-consciousness or self-understanding is a matter of *thinking* about myself in a certain way or making myself into an object of reflection. To understand myself as up to something at which I can succeed or fail must instead belong to my very being. Here, finally, we come to the point where the phenomenological account of intentionality, ontic transcendence, calls for completion in an account of 'primal transcendence'.

5. Existence as Primal Transcendence: Responsiveness to the Normative

A thing can be experienced as something only if it is taken to *count* as such a thing; that is, only if it is experienced in light of some (often not fully explicit) norm of what such a thing is supposed to be. Without this normative moment things would lack 'being' (*Sein*) in Heidegger's sense; that is, intelligibility, meaning. Lask and Husserl understood this as the 'validity claim' in all intentional content, an aspect of intentionality which, on pain of infinite regress, can be identified neither with a mental entity nor with a real part of the object. For this reason, first-order inquiry into an entity will never reveal that entity's 'being'. As Heidegger put it, 'one cannot pack transcendence into an

⁵⁴ There is, of course, much debate over whether machines (thermostats, computers) or non-human animals (wasps, tigers, dogs) *can* be said to have intentional content, experience something as something. We cannot enter into this literature here, but my own understanding of what further condition must be fulfilled is sketched in the following section.

intuition.’⁵⁵ We have seen that Heidegger traces the normative or validity character of our conscious experience back to an existential context of abilities, skills, and practices. Nevertheless, Heidegger also insists that being, meaning, cannot be phenomenologically clarified merely by substituting a ‘practical’ subject for the ‘theoretical’ subject of traditional transcendental philosophy. It may be that ‘one cannot pack transcendence into an intuition’, but ‘even less can it be packed into a practical comportment, be it in an instrumental-utilitarian sense or any other . . . Transcendence precedes every possible mode of activity in general, prior to *noesis*, but also prior to *orexis*.’⁵⁶ What makes our practical engagement transcendently constitutive of meaning is not *simply* the fact that it is efficacious or goal-directed. The instrumental rationality inherent in practices can provide the norms that condition intentional content only because it, in turn, is supported by ‘primal transcendence’ as the ‘ground for . . . every kind of ontic reason-for’.⁵⁷ What I encounter in the world can be held up to norms or standards only because in my very being I must hold myself to standards, that is, understand myself as *being* something that can succeed or fail. This primal transcendence—responsiveness to the normative *as* normative—is the ultimate ground of all transcendental conditions of possibility, or as Heidegger says, ‘the origin of “possibility” as such’.⁵⁸

Here I can develop this point only briefly.⁵⁹ Practical engagement can yield intentional content only if the normative conditions that make it the specific practice it is can be *experienced* as such conditions, that is, only if I can respond to them *as norms*, act not just in accord with them but in light of them. The in-order-to involvements of things depend for their disclosure on the norms inherent in abilities, skills, and practices. But a normative moment inhabits abilities, skills, and practices only because they ‘belong’ to a being who can acknowledge them *as* such conditions. Failing that, abilities, skills, and practices are underdetermined, that is, they cannot be understood as processes *governed* by norms unless one tacitly trades on how they would be experienced *by* such a being. In Heideggerian terms, they are what they are only within a ‘project’, an understanding of my *own* being as subject to normative evaluation.

Following Aristotle, Heidegger terms this further condition the ‘for-the-sake-of’ (*Worumwillen*) and identifies it with Dasein’s ‘understanding of [its own] being’. The instrumentally structured ‘totality of involvements itself goes back ultimately to a “towards-which” in which there is *no* further involvement’—that is, to something that is not defined by the instrumental nexus. But ‘the primary “towards-which” is the

⁵⁵ Heidegger, *Metaphysical Foundations of Logic*, 183.

⁵⁶ Heidegger, *Metaphysical Foundations of Logic*, 183. On p. 184 of this text Heidegger underscores that even ‘when theoretical comportment is apparently supplanted by the practical (primacy of practical reason) . . . the ancient approach remained directive’—i.e. ‘the genuine phenomenon of transcendence’ is ‘localized in a particular activity’. Hence Heidegger’s turn toward existential phenomenology is not equivalent to the ‘pragmatic’ turn in transcendental philosophy advocated e.g. by Karl-Otto Apel, Jay Rosenberg, and others.

⁵⁷ Heidegger, *Metaphysical Foundations of Logic*, 191.

⁵⁸ Heidegger, *Metaphysical Foundations of Logic*, 189.

⁵⁹ For further discussion see Steven Crowell, *Normativity and Phenomenology in Husserl and Heidegger* (Cambridge: Cambridge University Press, 2013).

“for-the-sake-of-which”, and ‘the “for-the-sake-of” always pertains to the being of *Dasein*, for which, in its being, that very being is essentially an *issue*’.⁶⁰ For instance, the practices of being a writer (the long hours at the computer, the wrangling with publishers, etc.) have salience for me—their demands take on normative force—only because I am *trying* to be a writer. To try is not merely to act in accord with norms (mechanically, as it were) but to be responsive to the normative, to the possibility of living up to the demands of what it is to be a writer or failing to do so. A monkey could perhaps try to write, but it could not try to *be* a writer; could not, in Heidegger’s terms, understand itself as a writer, act in light of writerly norms.⁶¹

In the phenomenon of the for-the-sake-of, or primal transcendence, we find Heidegger’s transformation of the transcendental subject, the ground of meaning or ‘world’ in Heidegger’s sense: ‘the basic characteristic of world whereby wholeness attains its specifically transcendental form of organization is the for-the-sake-of-which’.⁶² A particular project—understanding myself as a writer—can establish a particular totality of involvements in which things can show up as pens, paper, and the like, only because I am such that in my being that very being is an issue for me, is at stake for me; only because I *care* about what I am to be. As John Haugeland argues, the instrumental normativity of the in-order-to relations in light of which things are disclosed as suitable or useful can function as disclosive only so far as I am beholden to the ‘constitutive rules’ of a certain practice.⁶³ If I don’t care how file clerks are supposed to do things—if I do not try to act in light of the constitutive rules of clerkish practice—I cannot tell whether the various involvements of ‘this thing here’ make it a file cabinet; I can get no grip on what it is supposed to be, and so on the appropriateness or inappropriateness of its various aspects and properties. Being beholden to the rules of a practice is not a passive feature of my being, as though I were programmed to function in accord with them. Rather, it is to *commit* myself to *being* something. And committing oneself, trying, cannot be considered an act of consciousness. As Heidegger puts it: ‘In the projection of the for-the-sake-of as such, *Dasein* gives itself the primordial *commitment*. Freedom makes *Dasein* in the ground of its essence *responsible* to itself, or more exactly, gives itself the possibility of commitment’.⁶⁴ Transcendental subjectivity, ‘the being which is intentional’,⁶⁵ is not consciousness but commitment (*Bindung*).

Here, however, we must note a further crucial distinction, one indicated in Heidegger’s remark that the for-the-sake-of is that through which *Dasein* gives itself

⁶⁰ Heidegger, *Being and Time*, 116–17 (H84).

⁶¹ For some, this would not rule out attributing intentional content to the monkey. See Mark Okrent, *Rational Animals: The Teleological Roots of Intentionality* (Athens, OH: Ohio University Press, 2007).

⁶² Heidegger, *Metaphysical Foundations of Logic*, 185.

⁶³ John Haugeland, ‘Truth and Rule Following’, in *Having Thought* (Cambridge, MA: Harvard University Press, 1998), 305–61.

⁶⁴ Heidegger, *Metaphysical Foundations of Logic*, 192.

⁶⁵ Heidegger, *History of the Concept of Time*, 110.

the *possibility* of commitment. For commitment is always commitment to the norms of some *particular* way to be—and so to the norms of its constitutive practices—be it that of a writer, teacher, father, or whatever. But in order for me to be able to commit to particular norms I must already have understood myself in a normative light, grasped my existence in terms of the very *possibility* of distinguishing between better and worse, success and failure. Primal transcendence in this ‘ontological’ sense—transcendental subjectivity—is constituted by responsiveness to the normative *simpliciter*. Heidegger tries to suggest what this means by linking primal transcendence to Plato’s idea of the good, that is, ‘that on account of which something is or is not, is in this way or that’. The ‘*idea tou agathou*’, he writes, ‘which is even beyond beings and the realm of ideas, is the for-the-sake-of-which.’⁶⁶ To give oneself the *possibility* of commitment, then, is to understand oneself in normative terms, that is, in light of the good, of ‘what is best’. I can commit myself to being a writer, on this view, only because I have always already understood that to be a self at all is to be responsive to the normative—to the idea that what is can be held up to measure—and so responsible for the norms of this practice or that. And as I have tried to show, it is ultimately because I am ‘responsible’ in this way—bind myself to particular norms—that entities come to be held up to normative standards and thus come to be disclosed *as* something.

Thus phenomenology—all phenomenology—is transcendental to the extent that it provides an account of intentionality. This involves three stages. First, it recognizes the ubiquity of meaning and the priority of the question of meaning over metaphysical, scientific, or epistemic approaches; that is, it grasps the intentional object as ‘being in the sense of truth’. Second, it recognizes that meaning is not a conceptually or inferentially structured ‘logical space’ that can only be rationally reconstructed but the element of embodied practices and experience in which concrete ‘transcendental’ subjectivity is at home. Finally, it recognizes that the transcendental subject is neither a theoretical nor a practical one. The ‘being which is intentional’ must be understood transcendently not as a unity of apperception, nor as freedom in the sense of a practically rational agent, but as an entity whose being is defined by responsiveness to the normative as such, to the idea of the good.

⁶⁶ Heidegger, *Metaphysical Foundations of Logic*, 184.

12

Heidegger on Unconcealment and Correctness

Taylor Carman

Truth—together with *being* and *good*—is one of our most primitive concepts. It has also been, for that reason, one of the (more or less) perennial enigmas of philosophy. Nietzsche, for example, brooded and puzzled over whether and how it was possible to care about truth for its own sake. Of what value is truth itself, he wondered, once we see through all the old myths that dressed it up as something supernatural, transcendent, magical, and redemptive. Thus in *Beyond Good and Evil* he writes,

These are beautiful, glittering, tinkling, festive words: honesty, love of truth, love of wisdom, self-sacrifice for knowledge, heroism of truthfulness—there is something in them that makes one swell with pride. But we hermits and marmots, we convinced ourselves long ago in all the secrecy of a hermit's conscience that even this dignified pageantry of words belongs to the old false finery, junk, and gold dust of unconscious human vanity, and that the terrible underlying primary text *homo natura* must also be recognized beneath such flattering colors and painted surfaces. To translate man back into nature . . . Why would we choose it, this insane task? Or put otherwise: 'Why knowledge at all?'—Everyone will ask us this. And we, pressed to such a degree, we who have already asked ourselves the same question a hundred times, we have found and find no better answer.¹

Far from denying that truth has a value all its own, irreducible to any other value—power, justice, solidarity—Nietzsche evidently remained perplexed and anxious about his own commitment to truth precisely because of the tenaciousness of that commitment. But what makes such a commitment possible? Why should we care about truth at all? Even after setting aside all the obfuscating rationalizations of the moral-metaphysical tradition, Nietzsche confesses, 'we have found and find no better answer'.

¹ Nietzsche, *Beyond Good and Evil*, §230 (my translation); *Kritische Studienausgabe*, 2nd edn, ed. Giorgio Colli and Mazzino Montinari (Berlin: de Gruyter, 1988), v. 169–70.

Truth was also one of the central abiding concerns of Heidegger's thinking, early and late. What distinguishes Heidegger's reflections on truth from Nietzsche's, though, is precisely his attempt to describe its foundational status relative to the myriad other goods we more casually refer to as 'values'.² For Heidegger, by contrast, the Greek *alêtheia* 'is a word for what man wants and seeks in the ground of his essence, a word thus for something fundamental and ultimate (*Erstes and Letztes*) . . . what constitutes the ground and soil, the vault housing (*Wölbung*) human existence (*Dasein*)' (*GA* 34: 12).³ In conceiving of truth as ontologically fundamental in this sense, not as a mere value, but as a condition of the intelligibility of a whole range of practical and theoretical attitudes, Heidegger's account, unlike Nietzsche's, remains rooted in the tradition of transcendental philosophy.

The *locus classicus* of Heidegger's account of truth is §44, the final section of Division One of *Being and Time*. There he sets out to describe what he calls the 'ontological foundations' of the traditional concept of truth as agreement or 'correspondence' (*Übereinstimmung*).⁴ The ontological foundation of that concept, he suggests, is something more basic, namely 'uncovering' (*Entdecken*). Assertions can be true, that is, only if they are somehow involved in the uncovering of entities. Heidegger puts this, somewhat misleadingly, as a claim about what it *means* to say an assertion is true: 'The assertion *is true* means it uncovers the entity in itself. It asserts, it indicates, it "lets" the entity "be seen" (*apophansis*) in its uncoveredness. The *being-true* (*truth*) of an assertion must be understood as *being-uncovering*.'⁵

How do assertions manage to uncover entities? By figuring into human practices that do so. Being-uncovering is a possibility for assertions, that is, only because it is a possibility for human beings. As Heidegger puts it, 'Being-true as being-uncovering is . . . ontologically possible only on the basis of being-in-the-world. This phenomenon, which we have identified as a fundamental state of *Dasein*, is the *foundation* of the primordial phenomenon of truth' (*SZ* 219).

'Being true as being-uncovering is a mode of *Dasein*'s being . . . Uncovering is a mode of being of being-in-the-world' (*SZ* 220); 'only with the *disclosedness* (*Erschlossenheit*) of *Dasein* is the *most primordial* phenomenon of truth attained' (*SZ* 220–1). Assertions uncover, then, only because human beings do. This makes sense, since, after all, *asserting* is something people do.

Finally, Heidegger suggests, tracing the traditional concept of truth as correspondence back to its ontological foundations in the being-uncovering of assertions and

² Heidegger, 'Die Zeit des Weltbildes', *Holzwege*, 6th edn (Frankfurt: Klostermann, 1980), 100; 'The Age of the World Picture', *Off the Beaten Track*, ed. Julian Young and Kenneth Haynes (Cambridge: Cambridge University Press, 2002), 77.

³ Heidegger, *Vom Wesen der Wahrheit: Zu Platons Höhlengleichnis und Theätet. Gesamtausgabe*, xxxiv, 2nd edn, Freiburg lectures, winter 1931–2 (Frankfurt: Klostermann, 1997), 12. Hereafter *GA* 34.

⁴ *Übereinstimmung* is usually translated as 'agreement'. But since 'correspondence' is the standard English equivalent in philosophical discourse, I will use it instead.

⁵ Heidegger, *Sein und Zeit* (Tübingen: Niemeyer, 1927; 15th edn, 1979), 218. Hereafter *SZ*.

the disclosedness of Dasein, and so coming to see truth itself both *on the basis of* and *as* uncovering, reveals that the traditional concept misdescribes the phenomenon of truth, that truth is not, after all, any kind of correspondence between one entity and another: 'Truth thus by no means has the structure of a correspondence between cognition and object in the sense of a resemblance or alignment (*Angleichung*) of one entity (subject) with another (object)' (SZ 218–19).

In §44 of *Being and Time*, then, Heidegger advances (at least) four distinct claims:

- (1) that the traditional concept of truth as correspondence *presupposes* the phenomenon of uncovering;
- (2) that the truth of an assertion just *is* its uncovering an entity 'in itself' (*an ihm selbst*);
- (3) that the being-uncovering of assertions is possible only on the basis of Dasein's disclosedness or being-in-the-world;
- (4) that truth is *not* correspondence in the sense of a resemblance or alignment of one entity with another.

Many readers of *Being and Time* have thought it obvious that Heidegger rejects the notion of truth as correspondence altogether. But does he? He never says the notion is meaningless or incoherent. The worst he says about it in *Being and Time* is that it's 'very general and empty' (SZ 215), but of course even very general and empty notions can be coherent—abstract but benign, useless perhaps but harmless. At the beginning of his 1931–2 lectures, *The Essence of Truth*, he seems to go further by insisting that, although we ordinarily take it for granted as 'self-evident', the notion of truth as correspondence is in fact 'utterly obscure', 'ambiguous', 'unintelligible' (GA 34: 3, 4, 6).

Does that settle the matter? Well, not quite. When Heidegger says the received view of truth is 'unintelligible', what he means is not that it's demonstrably incoherent, but rather that we have no understanding of it:

Something is 'intelligible' to us if we understand (*verstehen*) it, i.e. can set ourselves before (*vor-stehen*) the thing, have its measure, survey and comprehend it in its basic structure. Is what we have just called 'self-evident' (truth as correspondence and correctness . . .) really *intelligible to us*? (GA 34: 2–3)

The answer is *no*. But again, this doesn't prove anything; it merely shows that we don't know what we're talking about when we say truth is correspondence. Heidegger says the notion of truth as correspondence is '*unverständlich*', but he also says in the same pages that it's merely '*unverstanden*', which can simply mean *misunderstood*. Could it be that all Heidegger wants to assert is that we do not yet have a proper understanding of the kind of correspondence that constitutes propositional truth? Could we come to understand it properly, and so embrace the traditional received notion, or at least a version of it?

Mark Wrathall has recently argued along these lines in what I think is the best brief discussion of Heidegger's account of truth. In support of his reading Wrathall cites

the following remarks from the beginning of *The Essence of Truth*. What makes an assertion true?

This, that in *what* it says, it corresponds to the things (*Sachen*) and states of affairs (*Sachverhalten*) *about which* it says something. The being-true of the assertion thus means such correspondence. What then is truth? Truth is *correspondence*. Such correspondence obtains because the assertion is directed to (*richtet sich nach*) that *about which* it says something. Truth is *correctness* (*Richtigkeit*). Truth is thus *correspondence, grounded in correctness, of the assertion with the thing*. (GA 34: 2)

So far, so good. But Wrathall quotes this passage as if it represents Heidegger's own considered view, whereas in fact Heidegger is here merely articulating what he takes to be our common preconception concerning truth, something we ordinarily take for granted as self-evident. He is not endorsing the concept of correspondence; indeed, as we have seen, he immediately goes on to say that that concept is obscure, ambiguous, and unintelligible. He is merely setting up the discussion, just as he does in *Being and Time* when he writes, 'The analysis sets out from (*geht aus von*) the traditional concept of truth and attempts to lay bare its ontological foundations' (SZ 214). To 'set out from' a received view is not to embrace it, but simply to take it as given, as received, in order to ask how it manages to be intelligible at all, *if* it does.

On Wrathall's reading, when Heidegger complains that the notion of correspondence is obscure, ambiguous, and unintelligible, all he is really denying is its *self-evidentness*, its *obviousness*. The idea of correspondence is not nonsense; it has merely been misunderstood. More precisely, it is, as Heidegger himself says, 'ambiguous.' Thus, in §44 of *Being and Time*, Heidegger asks us to consider the following paradigm case:

Someone with his back turned to the wall makes the true assertion, 'The picture on the wall is hanging askew.' This assertion proves itself (*weist sich*) when the speaker, turning around, perceives the picture on the wall hanging askew. What is the sense of the confirmation of the assertion? Is a kind of correspondence of 'knowledge' or 'what is known' with the thing on the wall established? Yes and no, depending on how phenomenally appropriately we interpret the expression 'what is known'. (SZ 217)

That reply ('yes and no') seems to support Wrathall's claim that the notion of correspondence is merely ambiguous and misunderstood rather than incoherent. The idea of correspondence is by itself unproblematic, that is, given a proper understanding of *what we know* when we know something. If 'what is known'—by which Heidegger means not the object, but the *content* both of the assertion and of the perception or belief that 'confirms' it as true—is supposed to be a kind of mental representation (*Vorstellung*), then *no*, the truth of the assertion is not any kind of correspondence between the representation and the object. What alternative notion of content would allow us to say *yes* to the idea of correspondence? Wrathall writes,

An assertion or proposition is true when it corresponds with a state of affairs . . . But this correspondence or agreement, Heidegger argues, cannot be understood on a representational model

of language. He argues instead that correspondence exists when our orientation to the world allows what is to show itself in a particular way, and thus it can be understood as a bringing out of concealment.⁶

On Wrathall's account, then, truth is correspondence, but correspondence just turns out to be whatever relation we have to things when we uncover them by means of perception or assertion. This leaves no room for the traditional image of correspondence as an agreement or fit between two similar or matching entities—say, two halves of a torn paper or a lock and a key. Indeed, it seems to leave no room for any notion of correspondence, except one defined trivially in terms of uncovering. What remains, on this account, still deserving of the name 'correspondence'? Did Heidegger merely want to preserve the word, not the image itself?

Wrathall's interpretation, it seems to me, is wrong, but only because, unlike most, it manages to be half right. He is right to insist that Heidegger neither doubts nor denies that the truth of perceptions, beliefs, and assertions consists in their *correctness*. He is wrong, I think, in assuming that for Heidegger correctness is the same as *correspondence*. Heidegger often lumps the two together, but not always. When he does distinguish them, moreover, it becomes clear that correctness is the more primordial phenomenon, both historically and conceptually. Disentangling them, I shall argue, promises to bring into sharper focus not only Heidegger's philosophical commitments concerning truth, but also the critical stance he takes up against the metaphysical tradition.

Heidegger advances two distinct claims, I believe, though he often runs them together. They were the first two of the four I distinguished earlier, namely:

- (1) that the traditional *concept* of truth as correspondence presupposes the phenomenon of uncovering;
- (2) that the truth of an assertion just *is* its uncovering an entity 'in itself' (*an ihm selbst*).

What (2) says, more precisely, is that the *correctness* of an assertion consists in its uncovering an entity in itself. What (1) says is that the *concept* of correspondence, the image of two adjacent things either resembling each other or fitting together like the pieces of a puzzle, is made possible by the being-uncovering of assertions and the disclosedness of Dasein. Uncovering 'explains' the concept of correspondence not by justifying or validating it, then, but by showing how it was ever possible for us to have such a concept. Whereas uncovering *constitutes* correctness, it merely *motivates the idea of* correspondence. And of course we can have the one without the other. Indeed, how could we do otherwise? How could we do without the concept of correctness? And how does the image of agreement or correspondence shed any further light on that concept?

⁶ Wrathall, 'Unconcealment', in Hubert Dreyfus and Mark Wrathall (eds), *A Companion to Heidegger* (Oxford: Blackwell, 2005), 338.

Granted, Heidegger often seems simply to equate correctness and correspondence (e.g. GA 45: 16). You might have noticed that in the passage from *The Essence of Truth* I quoted, Heidegger asks, 'Is what we have just called "self-evident" (truth as correspondence and correctness . . .) really intelligible to us?' (GA 34: 3; emphasis modified). That seems to suggest that correctness is no more intelligible to us than correspondence; that what goes for one, goes for the other.

You also might have noticed, however, that in the next passage I quoted, from the previous page, Heidegger says, 'Truth is thus correspondence, *grounded in correctness*, of the assertion with the thing' (GA 34: 2; emphasis modified). Moreover, the entire remainder of the lecture course is devoted to a scrupulously detailed examination of Plato's *Republic* and *Theaetetus*, texts in which, Heidegger supposes, the word *alêtheia* itself ceased to mean unconcealment and came to mean correctness (*orthotês*). Heidegger was wrong about that putative semantic shift in the Greek language, but he was probably right that theoretical correctness as such first became metaphysically paradigmatic and unconditionally important in Plato, in contrast to the Homeric and Presocratic tradition. 'Ever since', Heidegger writes, 'there is a striving for "truth" in the sense of the correctness of the gaze and its orientation. Ever since, in all fundamental orientations toward entities, what becomes decisive is achieving a correct view of the ideas.'⁷ Moreover, as Heidegger says here and elsewhere, only in Aristotle do we find the first version of what would become the standard account of truth as correspondence (*homoiôsis*), or, as Heidegger puts it, a kind of resemblance or alignment (*Angleichung*) (SZ 214) between experience and, as Aristotle says, 'those things of which our experiences are the images (*homoiômata*)'.⁸

What Heidegger describes in the lectures, then, is not the origin of the correspondence theory of truth, but the far more significant dawning of the idea that the essence of truth is not unconcealment but correctness. Again, Heidegger was wrong that the word *alêtheia* itself acquired a new meaning in Plato, a meaning he thinks it didn't have for the Presocratics. He was arguably right, however, that Platonic philosophy represents a radical shift in our understanding of the essence of truth. For Plato, as for the subsequent tradition, including our own current scientific-technological culture, truth is correctness, nothing more or less. What is correctness?

Rightness or correctness (*Richtigkeit*) is arguably the most basic concept of truth explicit in ordinary understanding, and probably always has been. Knowledge has a 'direction' (*Richtung*); it is 'directed' (*gerichtet*) straight at its object. 'True' in this sense means *right*, which, like the German *recht*, originally meant, according to the *OED*, 'Straight; not bent, curved, or crooked in any way'; 'Direct, going straight towards its

⁷ Heidegger, 'Platons Lehre von der Wahrheit', in *Wegmarken*, 2nd edn (Frankfurt: Klostermann, 1978), 232; 'Plato's Doctrine of Truth', *Pathmarks*, ed. William McNeill (Cambridge: Cambridge University Press, 1998), 179.

⁸ Aristotle, *De Interpretatione*, tr. E. M. Edghill; *The Basic Works of Aristotle*, ed. Richard McKeon (New York: Random House, 1941), 16a6.

destination'. In Old English 'wrong' occurred only as a noun meaning an injustice, but in Middle English the adjective meant 'Having a crooked or curved course, form, or direction; twisted or bent in shape or contour; wry'; 'Marked by deviation; deflected'; 'Mis-shapen; deformed'. Hence the verb 'to true', which means 'to place, adjust, or shape accurately; to give the precise required form or position to; to make accurately or perfectly straight, level, round, smooth, sharp, etc. as required'.

This understanding of truth as rightness over against wrongness, straight and direct as opposed to crooked and deviant, is plainly not the same as the concept of agreement or correspondence. It is also evidently a much more ancient notion. As Paul Friedländer observed, *pace* Heidegger, Homer already understood truth as correctness and always used *alêtheiê* and *alêthês* in connection with 'verbs of assertion', the object of which was not the unhidden, as Heidegger would have it, but—if such terms were negative at all, which Friedländer doubts—something more like 'that which is not-crooked', in contrast to 'everything that disturbs, distorts, slants'.⁹ In any case, Heidegger is right that Plato makes the image of direction or rectitude explicit and canonical in the theory of forms. Thus, in the allegory of the cave Socrates says, of the prisoner released from his bonds and now no longer captivated by the shadows but looking into the light, 'that now—because he is a bit closer to what is, and is turned toward things that *are* more—he sees more correctly (*orthotera*)'.¹⁰ In the same spirit, Socrates says in the *Theaetetus* that in false judgment, 'like a bad archer, one shoots wide of the mark and misses'.¹¹

The priority of rectitude to correspondence becomes clear, too, when we consider how naturally at home the former is in the expression of normative as well as factual truths. For while it remains a fruitless conundrum what in the world normative truths could be said to correspond to, we seem to have no trouble at all understanding wrongness as a kind of deviation, crookedness, or deformity.

The historical claim Heidegger is entitled to, then, is not that the meaning of the word *alêtheia* changed in the fourth or fifth century BCE, but that the ideal of theoretical correctness acquired centrality and prestige, a new intellectual and cultural authority it evidently lacked in the pre-Classical period. What interests Heidegger is not the history of the correspondence theory of truth—which is, after all, a pretty dismal affair—but a far more momentous event, namely the emergence of the scientific-theoretical understanding of truth as correctness.

Of correctness, moreover, Heidegger says two things: that it *is* a kind of uncovering, and that it is possible only *on the basis of* other, more primordial forms of

⁹ Friedländer, *Plato: An Introduction*, tr. Hans Meyerhoff (New York: Harper, Bollingen, 1958), 223. Friedländer willingly adds 'or conceals' because his point is that *alêtheia* was not semantically secondary to or parasitic on concealment. My point is just the evident archaic emphasis on truth as direction or rectitude. I am not suggesting that the image of straightness was primary or exclusive; in Greek, apparently, as in English, ethical and legal notions of honesty and reliability were at least as fundamental. Of course, such ethical concepts are themselves also deeply bound up with images of straightness and direction.

¹⁰ Plato, *Republic*, tr. C. D. C. Reeve (Indianapolis: Hackett, 2004), 515d.

¹¹ Plato, *Theaetetus*, tr. John McDowell (Oxford: Clarendon Press, 1973), 194a.

unconcealment, ultimately on Dasein's disclosedness. These were the middle two of the four claims I distinguished earlier:

- (2) that the truth of an assertion just *is* its uncovering an entity 'in itself' (*an ihm selbst*);
- (3) that the being-uncovering of assertions is possible only on the basis of Dasein's disclosedness or being-in-the-world.

What does it mean to say that correctness is itself a kind of uncovering? Does that claim constitute a substantive theory of truth?

No. To see why, it is important to see Heidegger's account in its proper context, namely *phenomenology*. The immediate and decisive predecessor to Heidegger's discussion is Husserl's phenomenological description of truth as correspondence in Chapter 5 of the Sixth Logical Investigation, 'The Ideal of Adequation: Self-Evidence and Truth'. There Husserl describes an experience of the coincidence or 'convergence' (*Deckung*) of intending acts with 'fulfilling' intuitions, paradigmatically perception. 'The fulfilling function of perception', he argues, indicates an 'ideal of ultimate fulfillment' of intentions in intuitive 'self-evidence' (*Evidenz*), which constitutes complete 'adequation to the "thing itself"' (*die "Sache selbst"*). He writes,

when a representational intention has attained final fulfillment through this ideally complete perception, then the genuine *adaequatio rei et intellectus* has been established: *the objectivity is really 'present' or 'given' exactly as it is intended*; no partial intention is further implied that lacks fulfillment.

'Self-evidence itself is', he continues, 'the act of the most complete synthesis of convergence.'¹² Husserl's theory, then, is an analysis of 'adequation' or correspondence as the convergence or coincidence of the contents of signifying and fulfilling acts. Truth is the ideal point of complete coincidence of intentions and intuitions. Heidegger's example of turning around to see that the picture on the wall is indeed askew, just as one said it was, is thus in effect a concrete illustration of the gist of Husserl's theory.

But if this is a theory, what *kind* of theory is it, and what is it a theory *of*? Is it a version of the correspondence theory? Husserl's aim is to do justice to the intuition driving that theory, cash it out in experiential terms, and so render it phenomenologically respectable. But the account also bears a resemblance to Peirce's conception of truth as the ideal endpoint of inquiry, an ideal one can never claim to have arrived at once and for all, but a kind of asymptotic limit. Finally, one might reasonably insist that Husserl's account is really a version of the coherence theory, since 'adequation' turns out to be an internal relation among the contents of intentional acts, not an external link between consciousness and reality itself. The fact is, virtually all theories of truth, even those intended as alternatives to the correspondence theory, try to do justice to the intuitive

¹² Husserl, *Logische Untersuchungen* (Tübingen: Niemeyer, 1900–1; 6th edn 1980), ii/2 (§37), 116, 119, 118, 122; *Logical Investigations*, 2 vols, tr. J. N. Findlay (London: Routledge, 1970, 2001).

appeal of the idea of correspondence. Even the redundancy theory (which merely says, ‘*p* is true, if and only if *p*’) purports to preserve all that is worth preserving in the idea of correspondence. This is why a deflationist like Paul Horwich can agree ‘that truths correspond to reality’, while denying ‘that such correspondence is what truth essentially is.’¹³ Correspondence, on this view, reduces to something harmless, but trivial.

Is Husserl’s theory a substantive theory of truth? Not exactly. It is, after all, part of a phenomenology of conscious experience. It is therefore doomed as a theory of truth for the same reason pragmatist and coherence theories are doomed. Why? Because, at the end of the day, everything they say about what constitutes truth is consistent in principle with the belief or proposition in question in fact being false. Such theories are plausible only to the extent that they adopt the same modesty that Horwich professes in his minimalist account, which does not purport to be a theory of some mysterious property called ‘truth,’ but merely describes and analyses our use of the truth predicate. No more than that, Horwich insists, is either possible or necessary.¹⁴ So too, it seems to me, Husserl’s theory can only be an account of the kind of experience that *motivates* our calling a perception or a belief true. If Husserl intended it as more than that—well, add that to a long list of unfulfilled intentions.

Heidegger’s account of truth, it seems to me, is descriptive, hence nontheoretical, in the same way. The account is meant not to explain what makes assertions true, but to describe the experiences and practices that *warrant* our calling them true, even if we turn out to be wrong. Indeed, Heidegger’s account is even more obviously modest than Husserl’s, since he says nothing about approximation to an ideal limit of perfect convergence and self-evidence, but says simply, ‘The assertion *is true* means it uncovers the entity in itself. It asserts, it indicates, it “lets” the entity “be seen” (*apophansis*) in its uncoveredness. The *being-true* (*truth*) of an assertion must be understood as *being-uncovering*’ (SZ 218). Surely he cannot have meant that every assertion that uncovers a state of affairs is *eo ipso* true. And yet this is what Wrathall’s reading seems to imply. He writes, ‘An assertion most genuinely succeeds if it brings a state of affairs into unconcealment for thought.’¹⁵ Is uncovering therefore a sufficient condition for truth? If so, then an assertion’s falsity implies that it doesn’t uncover after all, however much it may *seem* to.

This consequence raises the specter of Ernst Tugendhat’s classic objection to Heidegger’s account of truth as unconcealment, namely, that it fails to distinguish truth from falsity.¹⁶ Granted, it is hard not to regard other people’s false beliefs as, so

¹³ Paul Horwich, *Truth*, 2nd edn (Oxford: Oxford University Press, 1998), 116. Thus, seemingly tautological formulations in Plato and Aristotle, construed by some as statements of the correspondence theory, can also be read, not implausibly, as anticipations of the redundancy theory. See Plato, *Cratylus* 385b2 and *Sophist* 263b; Aristotle, *Metaphysics* 1011b25.

¹⁴ Of course, Horwich’s minimalism is theoretically ambitious in its own way, since, in helping itself to propositions, it takes their contents for granted and so stands in need of a theory of meaning. See therefore Horwich, *Meaning* (Oxford: Oxford University Press, 1998).

¹⁵ Wrathall, ‘Unconcealment’, 345.

¹⁶ Ernst Tugendhat, *Der Wahrheitsbegriff bei Husserl und Heidegger* (Berlin: de Gruyter, 1967), 329 *et passim*; ‘Heidegger’s Idea of Truth’, in Christopher Macann (ed.), *Critical Heidegger* (London: Routledge, 1996), 233 *et passim*.

to speak, obstacles standing between them and the world, as somehow blocking their view of the way things really are. But surely there is also a sense in which false beliefs and assertions uncover things, that is, make them manifest, bring them to light—albeit as *other than* they are. Indeed, Heidegger himself says as much in *Being and Time* when he describes the obfuscating effect of falling (*Verfallen*) and *das Man*: ‘Entities are not completely hidden, but precisely uncovered, but at the same time obscured (*verstellt*); they show themselves—but in the mode of semblance (*Schein*)’ (SZ 222). So, semblance or ‘seeming’ (*Scheinen*), which Heidegger earlier defines as ‘what is (*Seiendes*) showing itself as what it is *not* in itself’ (SZ 28), is a mode of uncovering, a kind of truth. Indeed, ‘Appearance and semblance are themselves founded, in different ways, in the phenomenon’ (SZ 31), that is, ‘that which shows itself in itself’ (*das Sich-an-ihm-selbst-zeigende*) (SZ 28).

Similarly, in *The Essence of Truth* Heidegger insists not only that the prisoners in the cave regard the shadows of the puppets on the wall as ‘the true’—or as Heidegger has it, ‘the unconcealed’ (*to alêthes*) (GA 34: 24)—but that Plato himself does. What Socrates actually says is that ‘what the prisoners would take for the true (*to alêthes*) is nothing other than the shadows of those artifacts’ (515c, translation modified). On Heidegger’s reading, the prisoners are wrong to believe that ‘the true is nothing other than the shadows’, but they are not wrong to regard the shadows themselves as (part of) ‘the true’ (*to alêthes*). For if *alêthês* means unconcealed, then the shadows are indeed *alêtheis*—though of course, as we readers know, not to the same degree or clarity as objects outside the cave. This reading of the text may be dubious, but it is crucial to Heidegger’s argument. On his account, what the prisoners see is *to alêthes*, since it is unconcealed to them. Indeed, for Heidegger’s Plato,

from childhood on, man is already and in his nature set before the unconcealed . . . Even in this strange situation in the cave, man is . . . directed to what is *before* him: *to alêthes*. It belongs to being human—and this is in the allegory already from the beginning—to stand in the unconcealed, or as we say, in the true, in truth. Being human means, however unusual the situation may be, not only but among other things, comporting oneself to the unconcealed. (GA 34: 25)

What matters to Heidegger’s account of unconcealment, then—even here, where what is at issue is apparently true and false belief, correct and incorrect opinion—is the phenomenological standpoint of the prisoners themselves, not what we readers know about the objective wrongness or their view of things. Indeed, Heidegger worries that calling what the prisoners see ‘shadows’ already misdescribes what is unconcealed to them in its unconcealment:

The prisoners indeed *see* the shadows, but not *as* shadows of something. When we say the shadows are for them the unconcealed, that is ambiguous, and we have already at bottom said *too much*. We, who already survey the entire situation, refer to what they have before them as shadows . . . it lies in the essence of their existence that precisely *this* unconcealed, which they have before them, *suffices*—so much so, that they don’t even know *that* it suffices. They are given over to that which *immediately* confronts them. (GA 34: 26)

This suggests that the position of the prisoners is so impoverished, so ‘immediate’, that they do not even apprehend what is unconcealed *as* unconcealed. This is not to say that the shadows are *not* unconcealed to them, but that the contrast between concealment and unconcealment is itself concealed from them, so that unconcealment as such remains concealed from them.¹⁷

In any case, for Heidegger, unconcealment is as definitive of false beliefs and appearances as it is of the true and right manifestations of things as they are, of what shows itself ‘in itself’. Hence Tugendhat’s charge that Heidegger’s theory of truth fails, since it fails to distinguish the true from the false. Tugendhat is not wrong about this, and indeed Heidegger himself eventually conceded the point.¹⁸ The objection misses the mark, however, precisely because it assumes that Heidegger’s account is meant to be a theory of truth, that is, a theory of correctness. Wrathall sees that this is mistaken and insists that, in Heidegger, ‘unconcealment is not to be taken as a (re)definition of propositional truth’. But again, he goes on to say that for Heidegger, ‘An assertion most genuinely succeeds if it brings a state of affairs into unconcealment for thought.’¹⁹ That sounds an awful lot like an analysis of correctness, and moreover it seems to imply that failed assertions (which I assume means *false* assertions) fail by failing to bring things into unconcealment. But that is evidently not Heidegger’s view.

It might be tempting to suppose that both true and false assertions uncover, but to different degrees—false assertions less, true assertions more. In this spirit, one might think, in *The Essence of Truth* Heidegger distinguishes between obscurity and opacity. Darkness can fail to make things visible, he says, precisely because what it is *is* insufficient light. A brick wall, by contrast, blocks vision, but since it is not the sort of thing that can make visible, neither can it be said to fail to do so. As Heidegger says, ‘Only that which is capable of affording can deny’ (GA 34: 56). So, just by being in the business of uncovering, perhaps even false beliefs do at least a little bit of what true beliefs do more of, namely uncover entities.

But there are two reasons this cannot be right: one textual, the other systematic. The textual reason is Heidegger’s fascinating—and, I think, compelling—suggestion that false belief uncovers not less but *differently* than true belief. Immediately following the remark I just quoted, he says, ‘The dark denies visibility because it can also afford

¹⁷ These remarks echo what Heidegger says elsewhere about the ‘world poverty’ of animals. *Die Grundbegriffe der Metaphysik: Welt-Endlichkeit-Einsamkeit, Gesamtausgabe*, xxix/xxx, Freiburg lectures, winter 1929–30 (Frankfurt: Klostermann, 1983); *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude*, tr. William McNeill and Nicholas Walker (Bloomington, IN: Indiana University Press, 1995).

¹⁸ In the 1964 essay, ‘The End of Philosophy and the Task of Thinking’, Heidegger writes, ‘In any case, one thing is clear: the question of *alêtheia*, of unconcealment as such, is not the question of truth. For this reason, it was inadequate and hence misleading to call *alêtheia* in the sense of lighting or clearing (*Lichtung*) “truth”. . . The natural concept of truth does not mean unconcealment, not even in the philosophy of the Greeks’ (*Zur Sache des Denkens*, 4th edn, Tübingen: Niemayer, 2000, 77); cf. *Time and Being*, tr. Joan Stambaugh (New York: Harper & Row, 1972), 70.

¹⁹ Wrathall, ‘Unconcealment’, 340–1. Wrathall (rightly) criticizes Tugendhat’s approach in ‘Heidegger and Truth as Correspondence’, *International Journal of Philosophical Studies*, 7 (1999), 69–88.

vision: in the dark we see the stars' (GA 34: 56). Indeed, it is precisely the darkness of the cave that allows the prisoners to see the shadows on the wall; too much light would wash them out, just as daylight washes out the stars. What exactly do darkness and light stand for in this metaphor, which Heidegger here extends well beyond Plato's original text? It is hard to say exactly, but they cannot simply be synonymous with concealment and unconcealment, since the point is precisely that daylight conceals the stars while the dark of night uncovers them. I am not suggesting that light and darkness in this passage symbolize true and false belief, and nothing else. Nevertheless, false belief is *like* darkness inasmuch as it obstructs and denies the clearest, most correct view of things. And yet, like the darkness of the cave, it also brings things into unconcealment—not less than true belief, but in its own way.

Which brings me to the systematic reason it cannot be right to say simply that false beliefs or assertions uncover less than true ones do. Recall that Heidegger worries that to call the shadows in the cave 'shadows' is to misdescribe the position of the prisoners, who do not, indeed *cannot*, see them *as* mere shadows. They *are* mere shadows, but they do not show themselves to the prisoners as such. Similarly, although it is obvious to us, and to the prisoner who escapes, that those in the cave can see less and less well than those in daylight, it is wrong to say that they *merely* see less and less well. Rather, they see something different, and they see differently. This is not to relativize the two points of view. Cave vision is inferior to vision in daylight, even if that fact only becomes obvious in daylight. The point is rather that cave vision as such, like false belief and assertion, in spite of its objective inferiority, affords those who have it genuine access to the world. Cave vision is a kind of vision, after all, and as such presents itself to its owners not as degraded or inferior, but as transparent and revealing. Again, this is why in *Being and Time* Heidegger defines mere 'seeming' (*Scheinen*) as 'what is (*Seiendes*) showing itself as what it is *not* in itself' (SZ 28). Someone who *has* a false belief, or is committed to a false assertion, that is, understands and experiences the world *through* it, by means of it, in its light. This is consistent with those of us who know better recognizing how the attitude or utterance fails to uncover things as they really are—that is, as *we* know them to be, thanks to *our* beliefs, which we cannot regard as false, so long as they remain ours.

Simply put, mere degree or intensity of uncovering cannot serve as a criterion of correctness for the simple reason that uncovering is a phenomenological notion and correctness is not. Uncovering is something in principle manifest from the first-person point of view. The being-uncovering of my attitudes or utterances cannot be held hostage to a third-person standpoint unavailable to me, as the rightness of my actions and the correctness of my assertions can be. My actions can always turn out to have been wrong, and my assertions false, regardless of how they struck me when I performed them; rightness as such has no phenomenological criteria. As Wittgenstein says, 'An inner experience cannot show me that I *know* something.'²⁰ He could have dropped

²⁰ Wittgenstein, *On Certainty*, ed. G. E. M. Anscombe and G. H. von Wright (New York: Harper & Row, 1972), §569.

the word 'inner' and said simply that no experience at all, no *phenomenon*, can show me that I know something, for truth is not a function of the way things show up for me, but whether the way they show up for me is the way they *are*. Attitudes and utterances, by contrast, cannot wholly fail to uncover entities while seeming to do so from the first-person perspective, for again, *having* beliefs and being *committed* to assertions are precisely ways of encountering entities. They are, as Heidegger says, modes of being-in-the-world.

Which brings me, finally, to my main point. The problem facing Heidegger's account is this. How can the correctness of an assertion *consist in* its being-uncovering, given that both correct and incorrect assertions uncover? What is the relation between uncovering and correctness? On the one hand, uncovering cannot be a *sufficient* condition for correctness, since both correct and incorrect assertions uncover, though in different ways. On the other hand, to say that uncovering is merely a *necessary* condition for correctness tells us nothing about what specifically distinguishes the true from the false, that is, the correct from the incorrect.

The solution to the problem, I think, is to say that the being-uncovering of assertions is neither a necessary nor a sufficient condition for their actual correctness, but rather a necessary (but insufficient) condition for our *regarding* them as correct. What kind of condition? In a word, *commitment*. Our regarding assertions as true is possible only on the basis of *making* them, that is, committing ourselves to them in linguistic practice. In the analogous case of belief, commitment just means *having* the belief: you must have a belief in order to regard it as true, though you need not regard it as true in order to have it.²¹

The acute reader will have surmised I have been referring throughout to belief and assertion indifferently in order to sidestep the vexed question of how the two are related. In *Being and Time* Heidegger gives pride of place to assertion and hardly mentions belief, since he thinks cognition is grounded in the pragmatic context of being-in-the-world. Belief looks more conducive to my argument, however, since one apparently cannot regard a belief as true without actually *having* it, though one can regard a proposition as true without actually *asserting* it. For present purposes, though, I think the difference makes no difference. Put in more pragmatic terms, my claim is that, as a matter of general principle, the collective practice of *regarding* assertions as true is possible only on the basis of a collective practice of *making* assertions. In both cognitive and pragmatic domains, that is, *deeming* or *holding true* is parasitic on the more fundamental phenomenon of *commitment*, whether in belief or in speech.

²¹ Here I am denying what Donald Davidson calls, with tentative approval, 'a redundancy theory of belief', namely, 'that to believe that *p* is not to be distinguished from the belief that *p* is true', 'Thought and Talk', *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984), 170. Davidson invokes the theory to support his thesis that 'a creature cannot have thoughts unless it is an interpreter of the speech of another' (157). It seems to me, conversely, that the fact that animals and prelinguistic children do indeed have beliefs of some kind, however rudimentary, renders the theory untenable. One man's *modus ponens*, as they say.

Identifying the disclosive function of belief and assertion with our *commitment* to them—that is, our *making* the assertion, our *having* the belief—solves the problem of the relation between uncovering and correctness. For it allows us to say both that false beliefs do indeed uncover entities, namely for those who *have* them, and that the being-uncovering of a belief is bound up with its being true as opposed to false, since we can *regard* beliefs as true only because we can *have* them. Indeed, as G. E. Moore observed, it is impossible to have a belief one considers false, for to regard it as false is precisely no longer to *have* it. Consequently, although we can regard false beliefs as uncovering entities *to others*, inasmuch as they have those beliefs, we cannot regard them as revealing the world *to us*. To say that uncovering is a condition of correctness, then, is to say neither that only true beliefs uncover entities nor that both true and false beliefs do so indifferently, but that *having* beliefs is a condition for regarding beliefs as true. It would be getting the phenomena backwards to suppose that we have beliefs *because* we regard them as true, as if the *commitment* to the belief were the consequence of a prior judgment concerning its truth.

To say that uncovering is grounded in Dasein's disclosedness or being-in-the-world, then, is to indicate the primacy of the *first person* as an ontological structure making intelligible the very idea that perceptions, beliefs, and assertions reveal the world at all, to anyone. Beliefs and assertions uncover entities, that is, only because human beings disclose worlds. Beliefs uncover entities to us in virtue of our *having* them, just as assertions uncover entities in virtue of our *making* them. Having beliefs and making assertions are modes of being-in-the-world that allow such actions and attitudes to draw things out into the light, even in the cave of ignorance and error.

13

Transcendental versus Hermeneutic Phenomenology in *Being and Time*

Cristina Lafont

As is well known, *Being and Time* is structured by two conflicting philosophical tendencies. On the one hand, the book is introduced and justified as an extraordinarily ambitious transcendental project. As Heidegger explains in the *Introduction*, the principal task of philosophy is to answer the question of the meaning of being. Since Dasein is the only entity that has an understanding of being, providing this answer requires following a transcendental strategy: 'fundamental ontology . . . must be sought in the *existential analytic of Dasein*' (SZ 13). As a crucial part of the overall project, *Being and Time* therefore aims to show 'the essential structures of Dasein that are determinative for the character of its being' (SZ 17).¹ On the other hand, however, the existential analytic of Dasein focuses on the hermeneutics of a factual Dasein in its average everydayness. In other words, the project of providing a fundamental ontology through an existential analytic of Dasein is the attempt to follow a transcendental strategy without a transcendental subject. In fact, in *Being and Time* there is no systematic use of any of the key conceptual resources of prior transcendental philosophy (e.g. the transcendental/empirical distinction, the transcendental reduction, transcendental subjectivity).

Ever since *Being and Time* was published these conflicting tendencies have forced interpreters to decide whether this work should be seen primarily as a continuation of transcendental philosophy by other (certainly innovative) means, or as a definitive

¹ The overall project is to demonstrate that time is the horizon for any understanding of being by showing that the essential structures of Dasein that are determinative for the character of its being must be interpreted as modes of temporality. In that sense, it does not aim at a *complete* ontology of Dasein. However, it clearly aims to exhibit 'not just any accidental structures, but essential ones which, in every kind of being that factual Dasein may possess, persist as determinative for the character of its being' (SZ 17).

break with transcendental philosophy. Either way, the challenge for any attempt to clarify the precise methodological status of that work is to produce an interpretation able to integrate *both* tendencies. From a methodological point of view, by aiming to single out the essential structures of human existence, *Being and Time* exemplifies the ahistorical mode of philosophizing characteristic of transcendental philosophy. But, from the point of view of its substantive claims, it emphasizes the radical facticity, historicity, and situatedness of human existence. It thereby calls into question the very idea of an ahistorical (i.e. time-transcending and uniform) way of experiencing the world. If humans are essentially self-interpretative creatures² and interpretation is by its very nature contextual and perspectival,³ what *Being and Time* discovers to be necessarily and universally valid about us is precisely that our disclosedness is essentially factual⁴ and thus that there can be no absolute truth for us.⁵

In a recent article titled 'Facticity and Transcendental Philosophy',⁶ Steven Crowell focuses on this central tension of *Being and Time* and provides an interesting defense of the interpretation of this work as a continuation of transcendental phenomenology. In contradistinction to other such defenses, he does not play down the importance of the facticity and historicity of Dasein, but insists on the need to confront the challenge they pose head-on. He recognizes that the focus on 'facticity' or 'situated subjectivity' in *Being and Time* is a genuine achievement, but he rejects any interpretation of that move as a 'de-transcendentalization' that is supposed to undermine the plausibility of Husserlian transcendental phenomenology. He argues that such interpretations cannot give an account of the strong methodological assumptions without which the transcendental project of *Being and Time* as originally conceived would make no sense at all. However fashionable they may have become, these interpretations fail because of their inability to confront this side of the challenge.

In what follows, I would like to defend the hermeneutic reading of *Being and Time* that I have articulated elsewhere⁷ by focusing on the challenge Steven Crowell has rightly pointed out. Although the 'detranscendentalizing' interpretations of *Being and Time* that he discusses⁸ are somewhat different from my own, I share with them the impression that the hermeneutic transformation of transcendental philosophy Heidegger undertakes in *Being and Time* undermines the plausibility of traditional transcendental strategies, be it of a Husserlian or other neo-Kantian variety. Certainly, the 'detranscendentalization' that I see at work in that transformation does not amount to a naturalization of the transcendental, nor to any blurring of the distinction between

² SZ 312. ³ SZ 148–53. ⁴ SZ 221.

⁵ Heidegger 1934, 36. Earlier in his lectures entitled *Einführung in die phänomenologische Forschung* Heidegger already indicates that 'it could well be that the idea of absolute validity is meaningless' (GA 17: 96).

⁶ Crowell 2003.

⁷ Lafont 2000, 2005.

⁸ Crowell distinguishes between what he calls 'bald aestheticism' and 'hermeneutic narrativism' as two different kinds of 'detranscendentalizing' interpretations of *Being and Time*. In his view the interpretations of Richard Rorty and Alexander Nehamas are examples of the former and those of Hans-Georg Gadamer, Charles Taylor, David Carr, Alasdair McIntyre, and Charles Guignon are examples of the latter.

empirical and a priori knowledge or, in Heidegger's terminology, between ontic and ontological knowledge. In that respect, I agree with Crowell that there is no room for denying the strong commitment to a priorism that underlies the original project of *Being and Time*. I disagree, however, with his tacit assumption that once this is conceded there is also no room left for the claim that *Being and Time* poses a radical challenge to traditional transcendental philosophy. Here I think Crowell misses a very important third alternative that in my view constitutes the *differentia specifica* between the hermeneutic and the traditional varieties of transcendental philosophy. It is to a detailed clarification of that alternative that I now turn.

1. Heidegger's Hermeneutic Transformation of Transcendental Philosophy

One of the main difficulties for any comparison of hermeneutic and traditional transcendental philosophy is that it is hard to characterize the latter with any precision. Not only are there numerous (and mutually incompatible) accounts of what distinguishes transcendental philosophy from other types of philosophy, but there are even those who claim that in fact nothing distinguishes the former from the latter.⁹ In view of this difficulty, and given that for my present purposes I do not need to identify its precise boundaries, in what follows I will discuss transcendental philosophy only in terms of its main, broadly agreed upon features. In so doing, I will leave the question of what the right interpretation of those features may be and whether they belong exclusively to transcendental philosophy entirely open.

In the previously mentioned article, Crowell identifies two features that in his view characterize transcendental philosophy. According to him, to say that philosophy is 'transcendental' is to say that it is *cognitive* and *autonomous*:

it is cognitive in the sense that it strives to attain truth about something, to 'get it right'. It thus belongs to the class of philosophical positions that see themselves as theoretical or scientific in the broad sense of the German term *Wissenschaft*. But in contrast to other members of this class—let's call them 'naturalistic'—transcendental philosophy does not see itself as continuous with the empirical sciences . . . It is autonomous or 'self-grounding' in the sense that it does not borrow premises from other cognitive domains: history, physics, psychology, and so on. (Crowell 2003, 101)

I agree with Crowell that the commitment to truth as well as to a priorism (i.e. to a sharp dichotomy between a priori and empirical knowledge) are characteristic of transcendental philosophy in general and of the project of *Being and Time* in particular. It seems clear that these characteristics are also sufficient for Crowell's own argumentative purposes. Given that the two varieties of 'detranscendentalizing' interpretations

⁹ For a recent defense of that claim see Pihlström 2004.

of *Being and Time* he discusses ('hermeneutic narrativism' and 'bald aestheticism') fail to account for one or both of these characteristics, this suffices to show their inability to provide a plausible account of *Being and Time* as originally conceived.¹⁰ However, for more general purposes, it also seems clear that these characteristics are insufficient to distinguish transcendental philosophy from other types of philosophy. Although strong a priorism may distinguish transcendental philosophy from most naturalistic varieties of empiricism,¹¹ it does not distinguish it from any variety of rationalism.

The additional characteristic that immediately comes to mind in this context is transcendental idealism. In spite of being the most controversial (and hardest to define) characteristic, it can hardly be ignored when the nature of transcendental philosophy is at issue. For present purposes, a minimal, negative characterization of transcendental idealism as the opposition to *metaphysical realism* will suffice. In order for a philosophical project to be identified as a transcendental project it seems necessary (although certainly not sufficient) that it reject any straightforward metaphysical realism. The exact reasons for that rejection are part of what distinguishes transcendental projects from one another. Following Kant, we can refer to this minimal feature of transcendental idealism as the 'Copernican revolution', or, to use another Kantian formulation, the realization that the conditions for the possibility of experience are simultaneously conditions for the possibility of the objects of experience.¹²

Although there is as much disagreement among Heidegger scholars on the precise combination of idealism and realism put forward in *Being and Time*¹³ as there is among Kant scholars on the precise nature of Kant's transcendental idealism, the methodological commitment to the Copernican turn is a fixed point for any interpretation of *Being and Time*, given that, as Heidegger puts it, 'fundamental ontology . . . must be sought in the existential analytic of Dasein' (SZ 13).

Whatever the exact idealist implications of that thesis may be, it seems clear that the existential analytic of Dasein is supposed to provide an account of those structures of any possible human experience that have ontological implications as regards any possible understanding of the being of entities. However, if this is the case, so Crowell's objection goes, there is no room for 'detranscendentalization' in that project. Claiming

¹⁰ As Crowell convincingly shows, these interpretations sooner or later break with the attempt to give an integrated interpretation of the conflicting tendencies of that work and argue instead for the position that 'Heidegger *should* have held in the late 1920s, though he failed to shed completely the Husserlian transcendental language committing him to the idea that philosophy ought to provide a "better account" of something' (2003, 103).

¹¹ Logical empiricism comes to mind as an exception, for it does accept the dichotomy between a priori (i.e. analytic) and empirical knowledge, although only on the basis of a conventionalist interpretation of the former (as pure 'knowledge of meanings' or 'meaning postulates'). See Ayer 1936 and Carnap 1950.

¹² Cf. Kant's supreme principle of synthetic judgments (*KrV*, A158/B197). In *What is a Thing?* Heidegger makes the following remark about Kant's principle: 'Whoever understands this principle understands Kant's *Critique of Pure Reason*. Whoever understand this . . . masters a fundamental condition (*Grundstellung*) of our historical existence, which we can neither avoid, leap over, nor deny in any way' (GA 41: 186).

¹³ For some recent discussions see Blattner 1994, 1999, Carman 2002, 2003, Han-Pile 2003, Lafont 2003, 2007, and Philipse 1998.

that those essential structures of human experience determine any objects of possible experience is tantamount to claiming that they are necessary and universal structures shared by all human beings. And this is tantamount to claiming that these structures possess a strictly transcendental status that philosophy can reveal a priori, that is, once and for all.

I certainly agree that *Being and Time* must be understood as a transcendental project. Moreover, I think that Heidegger himself understood it as a radicalization of Kant's own transcendental project. In Heidegger's view, asking about the transcendental conditions of perceptual experience, as Kant did, is just a specific case of the more fundamental inquiry into the transcendental conditions of understanding anything as meaningful. Accordingly, the hermeneutic conditions¹⁴ of understanding are the ultimate transcendental conditions of any human experience whatsoever, of which perceptual experience (of physical objects) is just a particular case. In that sense, getting the hermeneutic conditions of human understanding right implies providing universally valid results, as Crowell rightly claims. However, this is only part of the story. For once we recognize these hermeneutic conditions for what they are, they turn out to be very different from traditional transcendental conditions. Thus, even if one grants strict universal validity to the philosophical claims of Heidegger's analytic of Dasein, which *arguendo* I am prepared to do, one still can and must recognize the detranscendentalized status of the specific (i.e. contingent and historically variable) conditions of possibility of human experience that this analysis reveals. Even if all human beings are subject to the same hermeneutic conditions it does not follow that they are subject to the same possibilities of experiencing entities. Our experience of entities is indeed always dependent on our prior projections of their being. This is the truth behind Kant's transcendental idealism, according to Heidegger. But the fact that those projections are ours by no means implies, as Kant thought, that they are one and the same for all human beings. For Heidegger, to recognize that these projections are ours is precisely to recognize that they are *not* the product of pure reason or of a transcendental subject, but of Dasein *as it is, in its facticity and historicity*.¹⁵ As a consequence of this

¹⁴ I borrow the term 'hermeneutic conditions' from Carman 2003. I agree with his methodological view of their role in *Being and Time* as well as with his proposal to understand them on analogy with Allison's use of 'epistemic conditions' for interpreting Kant's transcendental project. However, I disagree with Carman's claim that, in contradistinction to Allison's use of 'epistemic conditions' to clarify the nature of Kant's transcendental idealism, in the case of Heidegger, the appeal to 'hermeneutic conditions' is unrelated to any defense of transcendental idealism. In my view, Heidegger's account of the hermeneutic conditions of human understanding involves a defense of idealism, although one based exclusively on hermeneutic reasons. I also disagree with Carman's specific account of some 'hermeneutic conditions' such as discourse, which in his view is essentially prelinguistic. But these particular differences do not diminish my agreement with Carman's view of *Being and Time* as an attempt to articulate the hermeneutic conditions of human understanding in analogy to Kant's articulation of the epistemic conditions of human experience.

¹⁵ In a letter to Husserl, Heidegger characterizes this detranscendentalizing step of *Being and Time* in the following terms: 'What is the kind of being of that being in which "world" is constituted? That is the central problem of *Being and Time*—i.e., a fundamental ontology of Dasein. It must be shown that the kind of being of human Dasein is totally different from that of all other entities, and that *as what it is*, it harbors *precisely in itself* the possibility of transcendental constitution' (Heidegger 1927, 601ff.; my emphasis).

detranscendentalizing step, however, two problems open up that were absent in prior transcendental philosophy and which call into question the traditional equation of apriority with universal validity. As I will try to show in what follows, in contradistinction to traditional transcendental philosophy, the opposition to metaphysical realism characteristic of hermeneutics involves the acceptance of *conceptual pluralism* and a strong *incommensurability* thesis.

2. Heidegger's Hermeneutic Idealism

As Heidegger explains at the beginning of *Being and Time*, according to its general signification 'being' is 'that which determines entities as entities' (SZ 13). In other words, it is what individuates entities and distinguishes them from one another. This signification is uncontroversial to the extent that it does not rule in or out any specific ontological or metaphysical claim. From a metaphysical-realist perspective, for example, that which determines entities as entities is supposed to be something that belongs to those entities themselves, that is, some ontic structure or properties that these entities have and others do not. According to this view, entities are already individuated one way or another independently of any human contribution. Consequently, there is a way the world is in itself regardless of whether humans are able to know it or not.

If we take metaphysical realism as a yardstick for realism, we can infer that Heidegger defends some kind of idealism from his further explanation of the meaning of 'being'. In contrast with the first signification, the next that Heidegger offers does seem incompatible with metaphysical realism. As he explains, 'that which determines entities as entities' is 'that on the basis of which entities are always already understood' (SZ 13). This equation rules out the possibility of a mismatch between what determines entities as entities and our understanding of them, whereas according to metaphysical realism such a mismatch cannot be ruled out in principle. Even in its non-skeptical varieties, the most that can be claimed according to metaphysical realism is that our understanding of the being of entities may *coincide* with what determines those entities as entities. When it does, we have knowledge of them; otherwise we do not.

According to Heidegger, however, such a view is deeply confused. It is the result of a basic failure to grasp the ontological difference or, in other words, to realize that 'the being of entities "is" not itself an entity' (SZ 6).¹⁶ Once the ontological difference is properly understood, it becomes clear that 'there is being only in an understanding of being' (SZ 212). Consequently, that which determines entities as entities is our understanding of their being; without such an understanding *nothing* individuates entities as

¹⁶ As he explains it in *The Metaphysical Foundations of Logic*: 'Weil das Sein nicht ist und daher nie und nimmer etwas am Seienden Mit-seiendes ist, hat die Frage, was das Sein am An-sich-Seienden sei, gar keinen Sinn und kein Recht . . . Wir erkennen immer nur Seiendes, aber nie seiendes Sein. Dies wird erst deutlich von der Transcendenz und der ontologischen Differenz her' (GA 26: 195).

entities. It is for this reason that fundamental ontology must be sought in the existential analytic of Dasein.

In his lectures on Kant titled *Phenomenological Interpretations of Kant's 'Critique of Pure Reason'*, Heidegger identifies this insight into the ontological difference as the truth behind Kant's Copernican turn. Commenting on the problem of the synthetic a priori Heidegger remarks:

Briefly the problem is the following: How can understanding open up real principles about the possibility of things, i.e., how can the subject have in advance an understanding of the ontological constitution of the being of a being? Kant sees this correlation, one which we formulate in a more basic and radical manner by saying: *Beings are in no way accessible without an antecedent understanding of being*. This is to say that beings, which encounter us, must already be understood in advance in their ontological constitution. This understanding of the being of beings, this synthetic knowledge *a priori*, is crucial for every experience of beings. This is the only possible meaning of Kant's thesis, which is frequently misunderstood and which is called his Copernican revolution. (GA 25: 38)

Here I will not attempt the difficult task of determining the exact nature of Heidegger's incorporation of Kant's transcendental idealism in the ontological difference.¹⁷ For present purposes it suffices to indicate that Heidegger agrees with Kant that the individuation of entities must be traced back to Dasein's fore-structure of understanding and that Dasein's projections of the being of entities, in virtue of their ontological significance, inherit the status that Kant ascribes to synthetic a priori knowledge. As Heidegger puts it, they are '*prior* to all ontic experience, but precisely *for* it' (KPM 8–9). On both counts it seems clear that Heidegger's agreement with Kant rules out the meaningfulness of metaphysical realism. However, it is the differences between those approaches that is at issue in this context.

According to Heidegger, Dasein's projections of the being of entities possess the features traditionally associated with synthetic a priori status:

- (1) they are prior to all experience of entities, but determine all experience of those entities, and
- (2) they cannot be revised on the basis of experience, since no experience can contradict them.

However, Heidegger offers an innovative explanation of these features which traces them back to *the circle of understanding*. In particular, as I will try to show in what follows, he explains assumption (1) in terms of a hermeneutic constraint on meaningful concept use, namely, the assumption that *meaning determines reference*, and assumption (2) in terms of a hermeneutic fact about interpretation, namely, the *holistic* structure of understanding. The crucial challenge to traditional transcendental philosophy involved in this innovative account of the a priori becomes clear as soon as one realizes

¹⁷ I have analysed this issue in some detail in Lafont 2003, 2007.

that nothing in these hermeneutic conditions rules out the historical alterability of Dasein's projections. Here lies the crucial detranscendentalizing step of Heidegger's transformation of the Kantian notion of apriority into the hermeneutic notion of the *a priori perfect* (i.e. the 'always already'). That something is 'in each case' *a priori* within a particular projection of meaning by no means implies that it is universally valid across different projections. This is why, as I mentioned before, the opposition to metaphysical realism characteristic of hermeneutics involves the acceptance of *conceptual pluralism* and a strong *incommensurability* thesis. In my view, these two features of the hermeneutic approach give a precise meaning to the otherwise vague talk of 'detranscendentalization' in *Being and Time*. Focusing on these features is thus the key to providing a specific answer to Crowell's challenge. Since I have explained this view in detail elsewhere,¹⁸ in what follows I will just provide a brief summary of how Heidegger's account of the hermeneutic conditions of understanding leads to these two features.

3. The Fore-Structure of Understanding

As is well known, one of the key tasks of *Being and Time* is to question the validity of the subject-object schema characteristic of modern philosophy, wherein all human experience is modeled on our perception of physical objects. From that perspective, establishing the priority of understanding over perception is the key to the hermeneutic approach. In section 32 of *Being and Time*, Heidegger takes this step explicitly when he claims that 'any mere prepredicative seeing . . . is, in itself, something which already understands and interprets' (SZ 149).¹⁹

In the context of justifying this claim, Heidegger provides a genuinely hermeneutic explanation of the most important feature of synthetic *a priori* status, i.e. that a projection of the being of entities is 'prior to all ontic experience, but precisely for it'. As I mentioned before, he interprets it as a consequence of a general constraint on meaningful concept use. The argument is as follows:

The circumspective question as to what this particular available thing may be, receives the circumspectively interpretative answer that it is for such and such a purpose. If we tell what it is for, we are not simply designating something; but that which is designated is understood *as* that *as* which we are to take the thing in question . . . The 'as' makes up the structure of the explicitness of something that is understood. It constitutes the interpretation. In dealing with what is environmentally available by interpreting it circumspectively, we 'see' it as a table, a door, a carriage or a bridge . . . Any mere pre-predicative seeing of the available is, in itself, something which already understands and interprets. (SZ 149)²⁰

¹⁸ See Lafont 2005.

¹⁹ Heidegger explains the importance of that step as follows: 'by showing how all sight is grounded primarily in understanding . . . we have deprived pure intuition of its priority, which corresponds noetically to the priority of the occurrent in traditional ontology. "Intuition" and "thinking" are both derivatives of understanding, and already rather remote ones' (SZ 147).

²⁰ Given Heidegger's claim that entities can only be discovered by a prior projection of their being (SZ 362), it should be clear that this argument applies not only to available entities but to all kinds of entities. In

Here Heidegger questions the possibility of a neutral perception of entities precisely by questioning the possibility of a neutral designation of them. The argument can be made explicit as follows. In order to refer to some entities we must first identify them so that they can be distinguished from others. But this cannot be done unless the terms used to designate those entities provide an understanding of what distinguishes them from others, that is, unless they provide the resources to identify entities as *what* they are, that is, in their *being*. To the extent that it is meaningless to purport to refer to entities whose conditions of identity one cannot possibly indicate, our understanding of the being of entities must determine in advance which entities we are referring to, that is, *meaning must determine reference*.²¹ This hermeneutic constraint explains why understanding is necessarily projective. It also provides a more or less trivial justification for the specific feature traditionally ascribed to synthetic a priori status, namely, that the understanding of the being of entities *determines all experience* of those entities. Given that the prior understanding of the being of entities is what makes our experience an experience of some specific entities (rather than others), it determines what these entities are (for us), that is, it determines *as what* they are accessible to us.²²

Heidegger develops this hermeneutic account of a priori knowledge in the reminder of section 32 of *Being and Time*. He indicates that the traditional conception of a priori knowledge is entirely unsatisfactory, for it does not recognize its internal connection with the phenomenon of projection. In order to show that understanding is necessarily both projective and presuppositional, Heidegger appeals to a well-known feature of the holistic activity of textual interpretation: the *circle of understanding*. In order to understand the meaning of a text we need to understand the meaning of its parts. But we can only understand its parts by anticipating the meaning of the text as a whole. Thus, as Heidegger puts it, ‘any interpretation, which is to contribute understanding, must *already* have understood what is to be interpreted’ (SZ 152). Without a *projection* of meaning no activity of interpretation can get off the ground. But for this very same reason interpretation is always *presuppositional*: ‘interpretation is never a presuppositionless apprehending of something merely presented to us’ (SZ 150). Consequently, an analysis of the conditions of possibility of understanding must provide an answer

fact, the context in which Heidegger makes this claim in *Being and Time* is the discussion of occurrent entities that are discovered through the scientific projection of ‘Nature’.

²¹ In *Vom Wesen der Wahrheit*, Heidegger explains: ‘We wish to consider the *essence of truth*. “Truth”: what is that? The answer to the question “what is that?” brings us to the “essence” of a thing. “Table”: what is that? “Mountain”, “ocean”, “plant”; in each case the question “what is that?” asks about the “essence” of these things. We ask—and yet we already *know* them! Indeed, *must* we not already know them, in order afterward to ask, and even to give an answer, about *what* they are? . . . Clearly, we must *necessarily* already know the essence. For how otherwise could we know what we should provide when we are requested to name truths? . . . *We must already know what and how the thing is about which we speak*’ (GA 34: 1–2; my emphasis).

²² As Heidegger explains it in the *History of the Concept of Time*, ‘it is not so much that we see the objects and things but rather that we first talk about them. To put it more precisely: we do not say what we see, but rather the reverse, we see what *one says* about things’ (GA 20: 75).

to the question of where our anticipations or projections of meaning come from. To answer this question Heidegger distinguishes three elements of the fore-structure of understanding: *fore-having*, *fore-sight*, and *fore-conception*. 'Fore-having' (*Vorhabe*) refers to the prior intelligibility with which we have understood in advance what we want to interpret. 'Fore-sight' (*Vorsicht*) refers to the specific perspective or point of view, to the specific understandings of being that guide a thematic interpretation. Finally, 'fore-conception' (*Vorgriff*) refers to the specific conceptuality, the particular vocabulary that is at the disposal of the interpretation. According to this view, understanding is always relative to a particular context, perspective, and vocabulary that together constitute what Heidegger calls the 'hermeneutic situation' out of which understanding arises and which we cannot transcend at will.

Thus, according to this analysis, the fact that our experience is determined by a priori structures is not a consequence of its being constituted by some invariant set of conditions to be discovered once and for all, as Kant thought. It is simply a consequence of the circle of understanding, that is, of the fact that 'every interpretation . . . must *already* have understood what is to be interpreted' (SZ 152). It is for this hermeneutic reason that entities can only be discovered by a prior projection of their being. However, and here lies the challenge to the traditional conception, this is something that any projection of the being of entities does. Therefore, the fact that such projections are prior to all experience of entities and cannot be revised on the basis of the experience of those entities by no means indicates their universal validity, since these conditions apply only to those who happen to share such a historically contingent projection.

This analysis of the hermeneutic conditions of understanding has important implications. First of all, it provides specifically hermeneutic reasons for rejecting metaphysical realism. If human experience does not arise primarily through a mere perception of entities but through a prior understanding of the being of those entities and if understanding is essentially holistic and perspectival, it follows that, contrary to the assumptions of metaphysical realism, the individuation of entities can only be holistic and perspectival as well.

As Heidegger argues throughout *Being and Time*, it does not make sense to ask how and what entities are in themselves without a prior determination of which specific meaning of 'being in itself', that is, which understanding of being, we have in mind. 'Real' or 'in itself' are specification-dependent terms. In *The Basic Problems of Phenomenology* Heidegger explains in detail the specific sense of his claim that 'there is being only in an understanding of being' (SZ 212). As he argues, an understanding of being as occurrence makes the individuation of occurrent entities (i.e. physical particulars) possible. But for the same reason, an understanding of being as availability (that is, an understanding based on functional categories and relations) is necessary to individuate available entities (i.e. equipmental particulars). He explains this as follows:

Equipment is encountered always within an equipmental contexture. Each single piece of equipment carries this contexture along with it, and it is *this* equipment only with regard to

that contexture. The specific *thisness* of a piece of equipment, its *individuation* . . . is not determined primarily by space and time in the sense that it appears in a determinate space- and time-position. Instead, what determines a piece of equipment as an individual is its equip- mental character and equipmental contexture . . . A being of the nature of equipment is thus encountered as the being that it *is in itself* if and when we understand beforehand the follow- ing: functionality, functionality relations, functionality totality . . . [That is,] if we have already beforehand *projected this entity upon functionality relation*. (GA 24: 414–15; emphasis in the original)

It is in this sense that Heidegger claims in *Being and Time* that a prior understand- ing of being as availableness provides the basis on which available entities ‘can for the first time be discovered as they are “substantially” “in themselves”’ (SZ 88).²³ On the basis of our understanding of being as availableness we are perfectly able to dis- cover available things as they are ‘in themselves’. We can distinguish whether a piece of equipment, say, a hammer, is a real hammer or not, whether it is a hammer ‘in itself’ or just a fake hammer, precisely because (and to the extent that) we understand in advance the criterion of appropriateness for that kind of available entity. Given that a hammer is for hammering, a hammer made out of dough, say, is not a ‘real’ ham- mer. Thus the hermeneutic discovery of the projective character of understanding questions the meaningfulness of the attempt to use terms such as ‘real’ or ‘in itself’ in an absolute sense, that is, independently of establishing in advance a criterion of identity or appropriateness that confers to them a determinate sense. For this rea- son, Heidegger’s claim that ‘availableness’ is the way of being of equipment in itself does not mean to exclude that these entities are also ‘occurrent’ (or as he puts it, that we can discover ‘something occurrent in what is available’: SZ 158). Equally so, he never denies that human beings are also ‘occurrent’ entities. What he does mean to exclude is the *reductionist* view that would claim that such entities are ‘really’ occur- rent entities, physical objects ‘in themselves’, and only available (or existent) in a ‘subjective’ sense. The scientific understanding of being (as occurrentness) is as con- textual and perspectival as any understanding always is. Within the parameters of its own fore-having, fore-sight, and fore-conception it is a perfectly acceptable kind of interpretation. What is unacceptable is its invasive attempt to monopolize the right to define reality in general and human reality in particular. It is in this sense that the projective view of interpretation leads to *conceptual pluralism*, that is, to the claim that there are many mutually irreducible interpretations of reality.

4. Incommensurability of the A Priori

This hermeneutic discovery has very challenging consequences for the traditional conception of a priori knowledge. It shows that what Kant erroneously thought were

²³ For a more detailed explanation see GA 24: 292–3.

the invariant features of any human experience whatsoever (i.e. the pure forms of intuition and the categories) are just a special case of what is in fact a much broader phenomenon, namely, the necessarily circular (i.e. temporal) structure of all human understanding. As a consequence, while Kant considered the special status of a priori knowledge to be due to the (alleged) fact that no human experience would be possible without that knowledge, for Heidegger the fact that scientific knowledge is based on an understanding of being as occurrence (and its corresponding concepts such as space, time, motion, force, etc.) by no means guarantees its absolute validity. It simply reveals the contingent fore-having, fore-sight, and fore-conception on which such knowledge is based.

In section 69 of *Being and Time*, Heidegger illustrates the historicity of the scientific projection of the being of entities by analyzing the historical transformation of the ancient conception of nature into the modern scientific one. In his opinion, the key to this transformation lies precisely in a change of 'projection' or, as it is called these days, a paradigm shift. In an astonishing anticipation of Thomas Kuhn's account of scientific revolutions, Heidegger explains that the ancient to modern shift does not consist merely in an increasing emphasis on observation or experimentation, but on the projection of an entirely different understanding of the being of entities. That is, it depends upon a new world-disclosure brought about through the establishment and definition of new basic concepts by modern scientists such as Galileo and Newton. To the extent that these new concepts organize all possible experience in advance, the grounding postulates or axioms of these modern theories, through which these concepts are defined, are at the same time responsible for the constitution of objects. To this extent, they have the status of synthetic a priori knowledge in the traditional sense. But, of course, this is only the case for those who share that projection. In fact, as Heidegger argues in *What is a Thing?*, the kind of synthetic a priori knowledge expressed in Newton's First Law of Motion, for example, 'was up until the seventeenth century not at all self-evident. During the preceding fifteen hundred years it was not only unknown; rather, nature and entities in general were experienced in a way with respect to which this law would have been meaningless' (GA 41: 79). For this reason Heidegger claims in *Being and Time* that 'before Newton his laws were neither true nor false' (SZ 227), in spite of their a priori status afterwards.

Given that, according to Heidegger, entities are only accessible through a prior projection of their being, it follows that entities made accessible by genuinely different projections are, by definition, not the same entities. That is, an alternative projection is (by definition) a projection of *different objects* and thus incommensurable with the prior one. This consequence of Heidegger's account of the hermeneutic conditions of understanding allows him to provide a very different hermeneutic explanation of the other crucial feature traditionally ascribed to synthetic a priori knowledge, namely, its immunity to revision by experience or observation alone. Heidegger undermines the traditional inference from immunity to revision to universal validity by tracing this

immunity back to a hermeneutic fact about interpretation, namely, the *holistic* structure of understanding.

In *What is a Thing?* Heidegger justifies the immunity from revision based on experience that he ascribes to the basic principles and axioms of scientific theories by appeal to what these days is called confirmation holism (i.e. the underdetermination of theory choice by evidence). To this end, Heidegger uses the example of different explanations for 'one and the same fact' within both the Aristotelian and Galilean paradigms, namely, the fact that under normal conditions in the earth's field of gravitation, heavy bodies pass through a determinate distance faster than lighter bodies do. He comments: 'Both Galileo and his opponents saw the same "fact". But they made the same fact or the same happening visible to themselves in different ways, interpreted it in different ways. Indeed, what appeared to them in each case as the authentic *fact and truth* was something *different*' (GA 41: 91). From this incommensurability among different projections Heidegger infers the impossibility of interpreting their historical change as a process of rational revision based on experience. As Heidegger claims in *Basic Questions of Philosophy*: 'it is simply pointless to *measure* the Aristotelian doctrine of motion against that of *Galileo* with respect to results, *judging* the former as backward and the latter as advanced. For *in each case, nature means something completely different*' (GA 45: 52–3; my emphasis).

Under the assumption that meaning determines reference, it follows that what 'nature' in each case means determines that to which the respective theories refer. Therefore theories with entirely different conceptions of natural entities cannot be about the same entities. But only if they were would it make sense to think of one as a correction of the other. Consequently, a scientific projection cannot be disproved by a different one; at most, it can be put 'out of force' by a different stipulation of what and how things are. And conversely, from the point of view of an old projection, the new one cannot be seen as better or worse but only as meaningless.

As I have argued elsewhere,²⁴ I think that there are good hermeneutic reasons to reject Heidegger's incommensurability thesis. In my view, it is precisely because Heidegger does not follow his own detranscendentalizing strategy to its natural conclusion that he still lends a priori knowledge the strong sense of transcendental determination that after his own relativization we realize it cannot possess. Borrowing Mark Sacks's useful distinction between 'transcendental features' and 'transcendental constraints',²⁵ my impression is that Heidegger's hermeneutic phenomenology

²⁴ Lafont 2000, 2005, and 2007.

²⁵ Sacks 2003 introduces the distinction between 'transcendental constraints' and 'transcendental features' in order to differentiate between a strong and a weak sense of transcendental determination. In my view, his characterization of the latter fits perfectly Heidegger's approach. His distinction is as follows: 'Roughly, a *transcendental constraint* indicates a dependence of empirical possibilities on a non-empirical structure, say, the structure of anything that can count as a mind. Such constraints will determine non-empirical limits of possible forms of experience . . . A merely *transcendental feature*, on the other hand, is significantly weaker. Transcendental features indicate the limitations implicitly determined by a range of available practices: a range comprising all those practices to which further alternatives cannot be made intelligible to those

reveals weak transcendental features, but mistakes them for strong transcendental constraints. Heidegger interprets the ontological difference in too close analogy to the transcendental/empirical distinction and thus embraces a hermeneutic idealism that should have been undermined by his own insights concerning the holistic structure of understanding.²⁶ Nonetheless, I think that Heidegger has given us good hermeneutic reasons to accept the relativization of *a priori* knowledge to specific conceptual paradigms or theories. In the end, it is this relativization that undermines the plausibility of traditional transcendental projects.

Heidegger's analysis of the historical transformation of science aims to show how the limits between the meaningful and the meaningless, the conceivable and the unconceivable, are relative to our historical, plural, and contingent projections of meaning. This view seems plausible in light of a repeated historical experience: what for centuries was thought to be a *a priori* true according to one conceptual paradigm turned out to be false according to the next. Moreover, the mismatch seems to occur for just the reasons that Heidegger emphasizes, which is that the experiences made possible by the development of the new scientific projection were inaccessible within the framework of the prior projection of entities. That is, from the point of view of the prior projection, objects under the descriptions of the later projection were literally 'inconceivable'. What this hermeneutic experience reveals is that a contingent understanding of the being of entities establishes limits to what can be meaningful or conceivable within a given projection. The lack of an alternative way of accessing entities makes a given understanding of their being contextually a *a priori*²⁷ within that projection in the specific Heideggerian sense that it cannot be questioned from within (i.e. on the basis of the experience that it itself makes possible). Statements that express such understanding of being (definitions, axioms, etc.) do indeed have the special synthetic *a priori* status traditionally ascribed to them: they express the ontology of the theory and are to that extent immune from revision by experience or observation alone. However, this is not because they are universally valid. As history has repeatedly shown us, often they are immune from revision simply because revising them has to await an alternative conceptual paradigm.

engaged in them' (Sacks 2003, 213). To the extent that human practices, concerns, interests, etc., stand to vary with empirical contingencies, 'we can say that transcendental features indicate the limitations on what, at a time, can be envisaged as possible by us . . . and to which alternatives cannot be made intelligible as long as those features retain their transcendental status . . . Consequently, those transcendental features of what we can currently envisage are not constraints on what is possible. When it comes to such transcendental features, inconceivability is no guide to impossibility' (Sacks 2003, 213). For a full analysis of the distinction see Sacks 2003, 198–218.

²⁶ In my opinion, Heidegger's insight regarding the holistic structure of understanding should have led him not only to relativize *a priori* knowledge to specific conceptual paradigms, but also to question the core assumption behind his interpretation of the ontological difference, namely, that there is an absolute and permanent dichotomy between *a priori* and *a posteriori* knowledge (i.e. *ontic* and *ontological* knowledge) and their respective kinds of truth. However, Heidegger never questioned this traditional assumption. For one of his latest statements of this view see Heidegger 1951/2.

²⁷ I borrow this term from Putnam 1962 and 1994. For a detailed comparison between Heidegger's and Putnam's conception of the synthetic *a priori* see Lafont 2007.

Be that as it may, in the present context I should not pick and choose what to take from Heidegger's hermeneutic phenomenology. Doing that may give the impression that I have failed to address Crowell's challenge head-on. Instead of showing how it was possible for Heidegger to see the project of *Being and Time* as a coherent enterprise at the time, I would be arguing for how he should have conceived it. Facing Crowell's challenge requires identifying those features of hermeneutic phenomenology that are sufficiently incompatible with traditional transcendental philosophy to undermine the plausibility of the latter without being so strongly incompatible with it as to undermine the plausibility of Heidegger's own understanding of *Being and Time* as a transcendental project. In my view, it is in virtue of its acceptance of conceptual pluralism and of the incommensurability of different projections of meaning that the hermeneutic approach undermines the plausibility of traditional transcendental philosophy without ever abandoning the horizon of a transcendental project. As I have tried to show, to the extent that our projections of meaning are responsible for our experience of entities, the detranscendentalization of the a priori entailed in Heidegger's hermeneutics does not amount to a rejection of transcendental philosophy. To the contrary, the opposition to metaphysical realism characteristic of transcendental projects is an essential component of the hermeneutic defense of conceptual pluralism. However, to the extent that our projections are historical, plural, and mutually inaccessible, such detranscendentalization amounts to a rejection of the traditional ascription of universal validity to them, and thus to the very attempt to single out any one of them as valid once and for all.

References

- Allison, Henry (1983), *Kant's Transcendental Idealism* (New Haven: Yale University Press).
- Ayer, A. J. (1936), 'The A Priori', in *Language, Truth and Logic* (Dover Publications: New York, 1946), 71–87.
- Blattner, William (1994), 'Is Heidegger a Kantian Idealist?', *Inquiry*, 37, 185–201.
- Blattner, William (1999), *Heidegger's Temporal Idealism* (Cambridge: Cambridge University Press).
- Carman, Taylor (2002), 'Was Heidegger a Linguistic Idealist?', *Inquiry*, 45, 205–16.
- Carman, Taylor (2003), *Heidegger's Analytic: Interpretation, Discourse, and Authenticity in Being and Time* (Cambridge: Cambridge University Press).
- Carnap, Rudolf (1950), 'Empiricism, Semantics, and Ontology', in *Meaning and Necessity* (Chicago: University of Chicago Press, 1958), 205–21.
- Crowell, Steven (2003), 'Facticity and Transcendental Philosophy', in Jeff Malpas (ed.), *From Kant to Davidson: Philosophy and the Idea of the Transcendental* (London: Routledge).
- Han-Pile, Béatrice (2003), 'Early Heidegger's Appropriation of Kant', in Hubert Dreyfus and Mark Wrathall (eds), *A Companion to Heidegger* (Oxford: Basil Blackwell), 81–101.
- Heidegger, Martin (1927), *Brief an Husserl*, *Husserliana* 9 (The Hague: M. Nijhoff), 600–2.
- Heidegger, Martin (1971), *Was heißt Denken?* (Tübingen: Max Niemeyer). [1951/2]
- Heidegger, Martin (1975), *Die Grundprobleme der Phänomenologie*, Marburger Vorlesung Sommersemester 1927, *Gesamtasgabe*, xxiv (Frankfurt: V. Klostermann, 1975). [GA 24]

- Heidegger, Martin (1977), *Phänomenologische Interpretation von Kants Kritik der reinen Vernunft*, *Gesamtausgabe*, xxv (Frankfurt: V. Klostermann). [GA 25]
- Heidegger, Martin (1984), *Die Frage nach dem Ding*, *Gesamtausgabe*, xxxxi (Frankfurt: V. Klostermann). [GA 41]
- Heidegger, Martin (1986), *Sein und Zeit* (Tübingen: Max Niemeyer). [SZ]
- Heidegger, Martin (1988), *Prolegomena zur Geschichte des Zeitbegriffs*, *Gesamtausgabe*, xx (Frankfurt: V. Klostermann). [GA 20]
- Heidegger, Martin (1988), *Vom Wesen der Wahrheit*, *Gesamtausgabe*, xxxiv (Frankfurt: V. Klostermann). [GA 34]
- Heidegger, Martin (1990), *Metaphysische Anfangsgründe der Logik im Ausgang von Leibniz*, Marburger Vorlesung Sommersemester 1928, *Gesamtausgabe*, xxvi (Frankfurt: V. Klostermann). [GA 26]
- Heidegger, Martin (1991), *Kant und das Problem der Metaphysik* (Frankfurt: V. Klostermann). [KPM]
- Heidegger, Martin (1991), *Logik*, Vorlesung Sommersemester 1934 (Madrid: Anthropos). [1934]
- Heidegger, Martin (1992) *Grundfragen der Philosophie*, *Gesamtausgabe*, xlv (Frankfurt: V. Klostermann). [GA 45]
- Heidegger, Martin (1994), *Einführung in die phänomenologische Forschung*, Marburger Vorlesung Wintersemester 1923/24, *Gesamtausgabe*, xvii (Frankfurt: V. Klostermann). [GA 17]
- Kant, Immanuel (1977), *Kritik der reinen Vernunft*, in *Werkausgabe*, iii, ed. Wilhelm Weischedel (Frankfurt: Suhrkamp). [KrV]
- Lafont, Cristina (2000), *Heidegger, Language, and World-Disclosure* (Cambridge: Cambridge University Press).
- Lafont, Cristina (2005), 'Hermeneutics', in Hubert Dreyfus and Mark Wrathall (eds), *A Companion to Heidegger* (Oxford: Basil Blackwell), 265–84.
- Lafont, Cristina (2007), 'Heidegger and the Synthetic A Priori', in Jeff Malpas and Steven Crowell (eds), *Transcendental Heidegger and Transcendental Philosophy* (Stanford, CA: Stanford University Press), 104–18.
- Philipse, Herman (1998), *Heidegger's Philosophy of Being* (Princeton: Princeton University Press).
- Pihlström, Sami (2004), 'Recent Reinterpretations of the Transcendental', *Inquiry*, 47, 289–314.
- Putnam, Hilary (1962), 'It Ain't Necessarily So', in *Mathematics, Matter and Method, Philosophical Papers*, i (Cambridge: Cambridge University Press, 1975), 237–49.
- Putnam, Hilary (1994), 'Rethinking Mathematical Necessity', in *Words and Life* (Cambridge, MA: Harvard University Press), 245–63.
- Sacks, Mark (2003), *Objectivity and Insight* (Oxford: Oxford University Press).

14

Merleau-Ponty's Transcendental Theory of Perception

Sebastian Gardner

I am concerned here with the status of Merleau-Ponty's theory of perception. Since my primary aim is to determine the *kind* of account offered by Merleau-Ponty, I will not offer a detailed discussion of Merleau-Ponty's highly original treatments of particular topics in the theory of perception, such as sensation, spatial awareness, or the role of the body. Instead I will argue that Merleau-Ponty's account of perception should not, in fact, be understood at all as a theory of perception in the usual sense of a theory formulated with a view to the solution of problems of psychological explanation and constrained accordingly; rather it should be understood as belonging to transcendental philosophy, conceived as a form of idealist metaphysics. If this is correct, evaluation of Merleau-Ponty's claims about perception needs to be cast in terms remote from those that a philosopher of mind applies to a theory of perception. Though I will not attempt here a full and final evaluation, I will set out what I take to be the basic justification offered by Merleau-Ponty for his transcendental claims.

There is a general issue regarding the relation of writings in the phenomenological tradition to analytic philosophy of mind. On the one hand it would seem that, whatever else it may comprehend, phenomenology is concerned in the first instance with the same topic as philosophy of mind: the phenomenologist is interested in mental states or phenomena and engaged, like the philosopher of mind, in making claims about their essential nature, necessary and sufficient or constitutive conditions, and so on. Accordingly it seems reasonable to expect that, allowing for differences of vocabulary and methodology, it will be possible to find points of convergence on matters of substance between phenomenology and philosophy of mind, and the recent literature on Husserl, Heidegger, Sartre, and Merleau-Ponty has suggested an abundance of these.

However, if what I will argue here is correct, then this view, for all its apparent plausibility, is mistaken with regard to Merleau-Ponty. Though nothing follows directly from

this regarding phenomenology in general, it does suggest a more general conclusion, namely that something essential to the phenomenological project necessarily goes out of focus in the attempt to read the phenomenologists as if their writings address the same questions as the philosophy of mind.

The chapter is organized as follows. In the first two sections I describe two competing interpretations of Merleau-Ponty. Section 1 outlines the view of those who affirm Merleau-Ponty's convergence with the philosophy of mind, which I will refer to as the Psychological Interpretation, and identifies considerations supporting it and some of its implications. Section 2 states briefly the Transcendental Interpretation, which views Merleau-Ponty in the light of the history of transcendental philosophy and claims to discover at the heart of his philosophical project an original form of idealism. The next two sections develop this interpretation. Section 3 considers how, on the Transcendental Interpretation, Merleau-Ponty's theory of perception and transcendental commitments are related logically. This, it will be seen, requires consideration of Merleau-Ponty's transcendental turn. Section 4 discusses Merleau-Ponty's use of the notion of ambiguity, since this, I argue, allows us to identify a clear line of descent from Kant and to grasp Merleau-Ponty's fundamental metaphysical thesis. Section 5 considers Merleau-Ponty's view of the relation of phenomenology and psychology, and his relation to the philosophy of mind.¹

1. The Psychological Interpretation

1.1

On what I will call the Psychological Interpretation, the *Phenomenology of Perception* attempts to establish certain claims regarding the nature of perceptual experience independently of any transcendental or metaphysical presuppositions.

The proponent of the Psychological Interpretation discovers in the *Phenomenology* a series of arguments for conclusions familiar from analytic philosophy of mind: against the concept of sensation, or a certain classical empiricist version thereof, and against the identification of perception with judgement or characterization of perceptual content as conceptual; and in favour of a rich and holistic theory of perceptual content which forges a deep constitutive link of perception with bodily capacities. Merleau-Ponty is interpreted as arguing on the basis of a familiar mixture of considerations of explanatory scope and completeness, theoretical perspicuity, fulfilment of epistemological desiderata, and phenomenological accuracy, his strategy being to measure philosophical theories of perception against our pre-philosophical concept of

¹ I concentrate throughout on the *Phenomenology of Perception*, and refer to earlier and later writings for supporting considerations. Page references to the English translation of the *Phenomenology* are prefaced PP; some quotations have been amended in light of the original. *PrP* refers to 'The Primacy of Perception and its Philosophical Consequences', *SB* to *The Structure of Behaviour*, and *VI* to *The Visible and the Invisible*.

perception, and to ask if the theories are faithful to the character that perceptual experience, in its full range, has for us.

The Psychological Interpretation is not obliged to deny that the *Phenomenology* contains metaphysical claims, but it will recommend that we attempt to understand these in the first instance as extrapolations from its prior, non-metaphysical claims about perception,² and if they cannot be so understood, then it will hold them to one side. The essential point for the Psychological Interpretation is simply the independence of the theory of perception, with respect to both the sense of its claims and the arguments given for them, from whatever metaphysics Merleau-Ponty may wish also to advance.

1.2

A number of considerations may be taken to support this view. It is in the first place suggested by the text of the *Phenomenology* itself regarding the content and order of its four divisions, the first of which (the Introduction) examines theories of perception with close reference to a large quantity of empirical material, and the second of which (Part One) pursues the connection of perception with the body. Not until the concluding chapters (the final chapter in Part Two, and Part Three) does Merleau-Ponty turn to metaphysical issues—those specific to human beings, including intersubjectivity, freedom, and self-consciousness—and briefly to general epistemological issues of truth and objectivity.

The Psychological Interpretation is supported also by the continuity of the *Phenomenology* with Merleau-Ponty's earlier *The Structure of Behaviour*, which provides a close examination of neurophysiological and functional theories of the organism, and much of which reads as a study in the philosophy of psychology. The *Phenomenology* begins with an explicit commitment to the phenomenological method, but otherwise may seem a direct extension of the line of holist, anti-reductionist thought begun in *The Structure of Behaviour*. Even the commitment to the phenomenological method which distinguishes the *Phenomenology* from the earlier work need not be regarded as signalling a real change of direction, in so far as the alliance with Husserl announced in the preface can be interpreted as a renunciation of any metaphysical premises for philosophical enquiry, and it soon comes to seem in any case that Merleau-Ponty's version of the phenomenological method is loosely defined and incorporates little of Husserl's purism and conception of rigorous science.

There is, furthermore, the obvious contrast of Merleau-Ponty with the other phenomenologists: Merleau-Ponty pays close attention to psychological science, and to its detail, rather than just referring in wholly general, critical terms to the very idea of empirical psychology. Sartre's early writings on imagination are also informed by empirical psychology, but Sartre uses it to chiefly negative ends, and *Being and*

² Thus Eilan suggests that Merleau-Ponty intends his theory of perceptual content 'to yield frank idealism' (1998, 353–4).

Nothingness begins with a statement of a set of supposed apodictic a priori truths concerning consciousness. By contrast, the *Phenomenology* seems to start on solidly a posteriori terrain: Merleau-Ponty appears willing to entertain, at least provisionally, the conceptual possibility that consciousness can be grasped in empirical-scientific, naturalistic terms.

The recent secondary literature—most prominently in the work of Hubert Dreyfus, Shaun Gallagher, and Sean Kelly—is rich with discussion of Merleau-Ponty exemplifying the Psychological Interpretation. It is argued that Merleau-Ponty contributes to contemporary debates in the philosophy of mind by providing arguments for the dependence of personal-level on subpersonal states, of conceptual on non-conceptual mental content, and of consciousness on embodiment; that Merleau-Ponty provides a convincing critique of the representationalism which holds sway in cognitive science and more generally furnishes insights which are obscured by cognitive science's bias in favour of cognition over performance; that Merleau-Ponty's account of skill acquisition stands in deep accord with developments in brain science neural network theory; and so forth.³ The overarching value of Merleau-Ponty for the philosophy of mind consists, on this view, in the fact that he denies the autonomy of the personal level of psychological explanation without any commitment to reduction to the physical, allowing his philosophy of perception to appeal both to those who argue for the necessity and integrity of the subpersonal domain opened up by cognitive science, and to those who favour a rich naturalism, who find in Merleau-Ponty a view of the mind which is non-materialist and non-reductionist yet also firmly anti-dualist.⁴

1.3

Among the consequences of recruiting Merleau-Ponty to the task of scientific investigation of the mental, and of claiming that his phenomenological studies support and receive support from empirical work, are (first) that his philosophical claims become subject in turn to empirical correction, and (second) that the task of explanation—in a *bona fide* and full-blooded sense, as opposed to the mere descriptive gathering and clarification of data—tends inevitably to pass out of the hands of phenomenology into neurophysiology and other more empirically tough-minded quarters. These consequences are however acknowledged and regarded as proper and acceptable by proponents of the Psychological Interpretation.⁵

Clearly the suggestion that phenomenology serves to clarify psychological *explananda* but lacks explanatory power of its own is worrying, and I will emphasize later that

³ A representative sample of relevant recent work includes Dreyfus 1999, 2005; Dreyfus and Dreyfus 1999; Gallagher 1998, 2008; Kelly 2002; see also Bermúdez 2005; Bickle and Ellis 2005, 158–9; Wider 1997, esp. 134ff.

⁴ See Gallagher 1998, 233, and Priest 1998, 3–4. Priest describes *The Structure of Behaviour* as a 'partial anticipation' of Strawson's *Individuals*.

⁵ See e.g. Dreyfus and Dreyfus 1999, 118; Gallagher 1998, 232–3; Kelly 2000, 162–5.

it runs contrary to Merleau-Ponty's intentions. For the present, another important implication of the Psychological Interpretation should be noted. The *Phenomenology* does not stop with a discussion of the nature of consciousness, experience, or mental content: as noted earlier, the concluding chapters of the work set out a general metaphysics of human existence. Moreover, the *Phenomenology* advances from its account of perception to a general metaphysical position which Merleau-Ponty wants to locate between idealism and realism, but which it is scarcely misleading to describe as idealist.⁶ In the relevant parts of the text it is quite clear that Merleau-Ponty's talk of perceptual experience as comprising 'pre-objective being', along with his critique of classical philosophical and psychological theories of perception as instances of 'objective thought', is fully metaphysical in intention. That is to say, talk of pre-objective being is not just talk of *experience* prior to the involvement of objectivity concepts in experience: it is talk of experienced *being* which is pre-objective.⁷ Nor is the critique of 'objective thought' equivalent to a critique of theories which deny the possibility of experience independent of objectivity concepts: it is also a critique of the *metaphysical* claim that objective representation is adequate to the representation of reality, or put the other way around, that reality is as objectivity concepts represent it as being. Pre-objective being and objective thought are, in Merleau-Ponty's full picture, terms of art belonging to metaphysics, not restricted to the philosophical analysis of mental phenomena.

Accordingly, the Psychological Interpretation is required to say of the *Phenomenology* that it contains a solid first argumentative half which establishes plausible conclusions regarding the nature of perception and the body, and a second argumentative half which, whatever its worth, lacks direct logical connection with the first.

The problem is that Merleau-Ponty seems unaware of this logical division, and this obliges the Psychological Interpretation to adopt a critical stance. There is no shortage of places, between adjacent sentences or within single sentences, where Merleau-Ponty must be regarded from the standpoint of the Psychological Interpretation as making a direct and unargued transition from philosophy of psychology to metaphysics. For example: Merleau-Ponty draws the conclusion regarding the body, from its possession of intentional properties and the asymmetry between how it is present to itself and how its objects are present to it, that the body is not in fact 'in' space at all, but rather 'inhabits' space (*PP* 139), and that an absolute, non-epistemological distinction must be drawn between the body qua object of science, the objective body, and the phenomenal body, the *corps propre* or *corps vécu*. Once we begin to look for them, such points—where Merleau-Ponty apparently confuses psychology-cum-epistemology with metaphysics, or distinctions of modes of presentation with distinctions of objects—are not hard to find.

⁶ I defend this classification in section 4.

⁷ Similarly, Merleau-Ponty's phrase 'the perceived world', *le monde perçu*, is not to be understood as equivalent to either 'our perception of the world' or 'the world, as we (happen to) perceive it', in the sense that one might talk about a mountain 'as seen from the south', implying that the world is one thing and our perceptual perspective on it another; 'perceived' and 'world' compose an indissoluble unity, the 'perceived-world'.

This point has been well made by Thomas Baldwin, who describes Merleau-Ponty's fundamental thesis that 'perception is "transcendental" in the sense that it cannot be adequately understood from within a fully objective, scientific conception of human life' as deriving from Merleau-Ponty's argument that 'because perceptual experience is epistemologically fundamental it cannot be the case that perception itself is fully comprehended within the explanatory perspective of natural science'.⁸ As Baldwin points out, if that is Merleau-Ponty's argument, then the naturalist will respond immediately that it rests on a confusion of epistemological with metaphysical priority; and instantly the whole anti-naturalistic, metaphysical aspect of Merleau-Ponty's philosophy appears to be the result of, to borrow a phrase of P. F. Strawson's from another context, a non sequitur of numbing grossness.⁹

The proponent of the Psychological Interpretation is thus driven to say that Merleau-Ponty sought a theory of perceptual content which avoids reducing it to either sensation or judgement, and that what he is right about, or at least offers a plausible defence of, is the idea that perceptual content is in a number of respects *sui generis*; his talk of the 'pre-objectivity' of perception should be translated into talk of the irreducibly holistic, non-conceptual, motor-conditioned, etc., nature of perceptual content. Merleau-Ponty's error was to think that, just because certain sorts of bad naturalistic theories of perception fail to do justice to the phenomena, naturalism itself must be rejected—he mistook the failure of *narrowly* empiricist theories of perception for the idea that perceptual experience cannot be a content of the natural world, or again, mistook the fact that perceptual experience lacks a certain very narrow sort of objectivity, for its non-objectivity *tout court*. To which it may be added that, had Merleau-Ponty been acquainted with the more sophisticated empirical psychological science of our present day, he might well have avoided this mistake.¹⁰

1.4

Doubts about the unity of the *Phenomenology* are not just a function of the Psychological Interpretation. The objection that the *Phenomenology* fails to hold together its philosophy of psychology with its metaphysics was put to Merleau-Ponty by the Hegelian philosopher Jean Hyppolite:

I would say simply that I do not see the necessary connection between the two parts of your paper [in which Merleau-Ponty had provided a synopsis of the *Phenomenology*—between the description of perception, which presupposes no ontology, and the philosophical conclusions which you draw, which do presuppose an ontology, namely, an ontology of meaning. In the first part of your paper you show that perception has a meaning, and in the second part you arrive at the very being of this meaning, which constitutes the essence of man. And the two parts do

⁸ Baldwin 1998.

⁹ See Bermúdez 2005, 299–303 and 315. See also Lacan's criticism, 1993, 74.

¹⁰ See Hammond et al. 1991, 278.

not seem to be completely interdependent. Your description of perception does not necessarily involve the philosophical conclusions of the second part of your paper. Would you accept such a separation? (*PrP* 39)

Merleau-Ponty's reply to Hyppolite's question was: 'Obviously not.' His immediately following statement was, however, perhaps not sufficiently full or clear to entirely silence Hyppolite's doubts: 'I have not, of course, said everything which it would be necessary to say on this subject. For example, I have not spoken of time or its role as foundation and basis' (*PrP* 39). One can see how temporality might provide some sort of bridge—the role played by temporality in Heidegger's *Being and Time* might for example be interpolated in the *Phenomenology*—but only traces of such an idea can be found in the *Phenomenology* itself,¹¹ and it is in any case hard to see how, even if it did restore the systematic unity alleged by Hyppolite to be wanting, temporality could also provide an effective basis for confuting the charge that Merleau-Ponty's general practice of transition from the theory of perception to metaphysics incorporates a fallacy. It remains to be shown, therefore, that the work is coherently unified in the way that Merleau-Ponty claims it to be.

2. The Transcendental Interpretation

2.1

What I will call the Transcendental Interpretation rejects the idea that the *Phenomenology* undertakes an enquiry into the nature of perceptual experience for its own sake: the purpose of Merleau-Ponty's enquiry into perception, it maintains, lies in its contribution to a transcendental theory with metaphysical implications.¹²

Merleau-Ponty provides in his discussions of vision and of the body early in Part One many statements of how the conditions that his phenomenology uncovers are intended to be in the true and genuine sense transcendental, i.e. a priori and necessary, and non-identical with empirical, contingent, or mundane states of affairs. He denies that the structure of vision, its perspectival articulation and figure/ground form, is due to 'the contingent aspects of my bodily make-up, for example the retinal structure' (*PP* 67–8). Similarly the permanent and ineliminable presence of the body—along with other of its features, including its affectivity—is described as a necessity that is not 'merely physical' but rather 'metaphysical' (*PP* 91). Kant's Copernican form of explanation is employed in the argument that Merleau-Ponty gives for this thesis, which corresponds closely to Kant's argument regarding space and time in the metaphysical

¹¹ Relevant remarks are scattered over *PP* 410–11, 425–6, 428–33.

¹² See Dillon 1998, ch. 2; the concluding chapter of Langer 1989; and Priest 1998, 224–5. Priest's broadly Kantian interpretation of Merleau-Ponty differs from mine by ascribing to him a non-idealistic, minimalist conception of the transcendental (see 80, 99, 254 n. 3); in consequence Priest locates inconsistencies and limitations in Merleau-Ponty (see 1998, 235, 238, 255 n. 11) which I think an idealistic interpretation avoids.

expositions of the Transcendental Aesthetic: the body's permanence cannot be 'a necessity of fact, since such necessity presupposes' it, and 'factual situations can only impact upon me if my nature is already such that there are factual situations for me' (PP 91). Merleau-Ponty affirms, therefore, the distinction of transcendental from empirical necessity,¹³ and that the subject's mode of cognition has Kantian explanatory priority over the objects of cognition.

2.2

The positions under attack in the *Phenomenology* are grouped under the general heading of 'objective thought', and fall into two kinds, each identified with a different form of philosophical explanation.¹⁴ Empiricism seeks to explain the objectual character of experience in terms of relations between an independent natural reality and human subjects conceived as items located within its causal nexus; Intellectualism treats the objectual character of experience as the product of subjective operations guided by thought. Empiricism includes various forms of classical empiricist philosophy, scientific realism and naturalism, while Intellectualism encompasses various forms of seventeenth-century rationalism, Kant, Husserl, and Sartre.¹⁵ Both are defined with reference to a particular, highly abstract, transcendental *explanandum*, namely the *objectual character* of experience, its articulation into objects and its character *as* experience, i.e. as involving a relation of subject to object. Empiricism deserves the label 'objective thought' because it takes as given the thought of a pre-articulated realm of *objects* (in which human subjects are included); Intellectualism does so because its explanatory bottom-line is provided by *thoughts* of objects.

The overall argument of the *Phenomenology* is designed accordingly to criticize the various theories of Empiricists and Intellectualists in a unified way which leads to the identification of a common underlying error, and to set out an alternative account of the objectual character of experience. The two aims of the *Phenomenology*—the negative, critical-diagnostic work, and the provision of a positive alternative—are of course not independent: the common error is the assumption of objective thought, to which Merleau-Ponty's own theory of transcendental conditions is meant to provide the only possible alternative. Merleau-Ponty's theory, in the briefest statement, is that the fundamental ground of the objectual character of experience lies in the pre-objectivity of perception: this, he argues, makes it possible for a subject to be presented with an

¹³ But not of course—as will become clear—Kant's view of the nature and grounds of transcendental necessity.

¹⁴ The Psychological Interpretation, by contrast, distinguishes these more narrowly as different forms of *psychological* explanation.

¹⁵ There are complications here, in so far as in none of the last three cases does Merleau-Ponty regard *all* aspects of Kant, Husserl, and Sartre, by any means, as Intellectualist. The classification of Husserl as an Intellectualist, in particular, raises questions. Plausibly, Merleau-Ponty's paradigm of Intellectualism is supplied largely by French neo-Kantianism. To keep matters simple I will not pursue these issues.

articulated realm of objects, and it also allows us to understand how reflection can be led astray into thinking that what makes this realm possible is instead either the objects themselves or our thoughts of them.

3. Theory of Perception and Transcendental Metaphysics

3.1

The key question for the Transcendental Interpretation concerns the logical relation of the *Phenomenology's* theory of perception and its transcendental metaphysics. There are three possibilities:

- (A) That the transcendental metaphysics logically *follows from* the theory of perception.
- (B) That the transcendental standpoint, from which the transcendental metaphysics will be developed, is assumed at the outset but only *provisionally*, as a hypothesis to be tested and confirmed by the discussion of perception.
- (C) That the transcendental standpoint is assumed from the outset *non-provisionally* by the discussion of perception.

I will argue that, although there are grounds for thinking that Merleau-Ponty's argumentative intentions are not fully clear, his considered view veers towards (C), which also makes the best sense of his position.

3.2

Let us begin with (A). If Merleau-Ponty's intention were to present in the *Phenomenology* a sequence of argumentative steps—first a refutation of naturalism, then a critique of Kant and Husserl, followed by an account of their common objectivist error, concluding with the correct transcendental theory—then the work as a whole could be regarded as avoiding any philosophically significant presuppositions, and so as arguing from scratch and in a linear manner for Merleau-Ponty's transcendental-metaphysical position.

This reading is attractive for an obvious reason: if the *Phenomenology* proceeds via an internal critique of naturalism or at least a critique on grounds which avoid transcendentalist presuppositions, to establish the general necessity for a transcendental approach, then the earlier chapters of the *Phenomenology* comprise an *argument for the transcendental turn*.

This is not, however, what we find. Consider the *Phenomenology's* arguments against Empiricism. Merleau-Ponty assembles numerous instances where Empiricist explanations are revealed to have gaps. This does not, however, spell an end to Empiricism, and indeed it is hard to see how it could do so, since every point of incompleteness in Empiricist explanation simply provides—in the eyes of the Empiricist—a new *empirical*

explanandum which stimulates the development of an improved empirical theory. For example, if 'sensation' as classically conceived does not facilitate the discovery of psychological laws, or otherwise impedes empirically significant theory, then what follows is just that scientific psychology should substitute a different conception of the original causal input to cognition. Merleau-Ponty could discredit this movement towards increasing sophistication in Empiricism only if he could show (a) that the gaps in extant empirical explanations are *in themselves not empirical*, or (b) that empirical explanations of perception are *intrinsically faulty*. But since, as noted previously, Merleau-Ponty does not seem to want to argue in Sartre's fashion that the very idea of treating the mental in terms of efficient causality is conceptually awry, the only route that he has to (b) would seem to be via (a), and it is hard to see what could *compel* the Empiricist to accept that an empirical gap is in truth a manifestation of non-empirical being. Merleau-Ponty himself is fully aware of this difficulty:

empiricism cannot be refuted . . . Generally speaking, the description of phenomena does not enable one to refute thought which fails to grasp itself and takes up residence in things [i.e. objective thought]. The physicist's atoms will always appear more real . . . The conversion of point of view must be undertaken by each one for himself, whereupon it will be seen to be justified by the abundance of phenomena which it elucidates. Before its discovery, these phenomena were inaccessible, and to the description of them which we offer empiricism can always retort that it *does not understand*. (PP 23)

If we examine the text of the Introduction and Part One in the light of these remarks, we find that at the crucial points where an argument for the transcendental turn might have been expected, Merleau-Ponty does indeed simply jump from the identification of a gap in empirical explanation to a transcendental assertion. Examples were given in section 2—of the body's not being 'in' space, and of the distinction between the objective and the phenomenal body—and many others can be found. Having argued in the chapter on sensation that no such unit of experience exists, Merleau-Ponty concludes that the concept of perception 'indicates a *direction* rather than a primitive function' (PP 12). In the chapter on association, having shown that bare association is unable to yield an analysis of memory, Merleau-Ponty asserts that one must therefore admit 'an original text which carries its meaning within itself . . . this original text is perception itself' (PP 21). Much of what Merleau-Ponty has to say in these places against Empiricism simply invokes, with some modification of terminology or emphasis, Kantian or Husserlian lines of thought, as if he were regarding the transcendental turn as a *fait accompli*, executed already and decisively earlier in the history of philosophy. But if this is so, then Merleau-Ponty is taking transcendental anti-naturalism for granted: the argument for it must be offstage in the *Phenomenology* and the Introduction's critique of Empiricist theories of perception regarded as presupposing arguments given already by Kant or Husserl.¹⁶ That Merleau-Ponty does not intend to argue to the transcendental position from scratch is, on the face of it, what he tells us

¹⁶ It might be conjectured that the transcendental turn which is offstage in the *Phenomenology* is regarded by Merleau-Ponty as having been effected in his earlier *The Structure of Behaviour*, and that this is supported

when in the preface he avows a commitment to phenomenology conceived as 'a study of essences', 'a transcendental philosophy', 'a rejection of science' (PP vii–viii).

Consider next Merleau-Ponty's account of the 'reduction to the pre-objective' in *The Visible and the Invisible*.¹⁷ Here Merleau-Ponty appears to want to introduce and justify the phenomenological method which will take us to his transcendental conclusions on the grounds that it (alone) is *presuppositionless*. The notion that philosophy should proceed from such a starting point recalls a string of modern philosophers from Descartes to Husserl, but a difficulty confronts the supposition that Merleau-Ponty is following that well-trodden path. The problem is simply that what Merleau-Ponty claims we discover when we discard our presuppositions is nothing less than the 'inverse' of common sense (VI 157). Common sense, he tells us, attempts to 'construct perception out of the perceived': it theorizes 'causes' of perception which act on us (VI 156) and thereby presupposes 'correlatives or counterparts of the *objective* world' (VI 157).¹⁸ It follows that Merleau-Ponty's presupposition-free realm of phenomena is *inaccessible* merely on the basis of a suspension of ordinary judgements of objects' empirical reality: access to the phenomena obscured by common sense requires a *positive*, purgative operation.¹⁹ Waiving the problem that on the face of it some theoretical apparatus is surely required for this operation, it must in any case have been shown beforehand that the common sense 'given' is *contaminated* with presuppositions, i.e. that what is given to common sense not merely *has* presuppositions but that those presuppositions are cognitively *defective*. And it cannot be a strictly epistemological motive that has led us to this conclusion, since Merleau-Ponty

by Merleau-Ponty's claim at SB 206 that what issues from the examination of scientific thought is a transcendental idealism. The problems with this suggestion are that *The Structure of Behaviour* does not contain a convincing argument for the transcendental position, and that Merleau-Ponty gives no sign of thinking that the *Phenomenology* relies on its argument (which would make little sense in view of the philosophical differences between the two works and the *Phenomenology*'s claim to be philosophically fundamental).

¹⁷ VI 156–60.

¹⁸ See PrP 19ff.; and SB 219: 'it is natural for consciousness to misunderstand itself precisely because it is a consciousness of things'. Note that there are differences of emphasis in Merleau-Ponty's accounts of the formation of objective thought. In his discussion of the body at PP 94–5, where Merleau-Ponty tries to determine what has led to objective thought's conception of the body—and thus, since the body is a central transcendental condition, to the general metaphysics of objective thought—the factors that he cites have all to do with the scientization of commonsense ways of thinking. But in other passages, such as the extremely clear account on PP 70–1, objective thought is ascribed to both science and common sense, but science plays no role at all in its *formation*, this being attributed entirely to the simple attainment of objectivity in perception, in the sense of taking things to have a subject-independent, non-perspectival existence and constitution: the 'positing of the object . . . makes us go beyond the limits of our actual experience' (PP 70), and leads to 'absolute positing' of the object, which is 'the death of consciousness, since it congeals the whole of existence' (PP 71). I take it, as section 4 of the chapter should make clear, that the second view represents Merleau-Ponty's true position.

¹⁹ Merleau-Ponty's distinction between pre-objective perception, and perception mediated by objective thought, has obvious affinities with, though does not strictly correspond to, Heidegger's distinction between authentic and inauthentic temporality, Sartre's distinction between consciousness and the degraded 'psychic' posited by impure reflection, as well as Husserl's original distinction between consciousness purified by the phenomenological *epoché* and consciousness interpreted according to the natural attitude.

evinces none of the relevant epistemic anxieties concerning objectivity and rationality; he shows, for instance, no interest in retracing the skeptical, certainty-orientated route of Descartes's *Meditations*.²⁰ Merleau-Ponty instead motivates his call to avoid presuppositions with the statement that 'the enigma of the brute world is finally left intact by science and by reflection' (VI 156), which may suggest (A), but in fact does not help us to construe the adoption of the transcendental standpoint as motivated sufficiently by a prior and independent critique of Empiricism: for the *relevant sense* in which science and reflection leave the enigma of the brute world 'intact', i.e. unexplained, and in which the bruteness of the world can be designated an 'enigma', cannot be grasped unless transcendental conceptions are already in play; as we saw Merleau-Ponty imply in the passage quoted from *PP* 23, some prior, alternative philosophical conception must be presupposed if Empiricism is to be grasped as having explanatory limitations.²¹

3.3

Before jumping to affirm (C), we should briefly consider (B), which might seem to accommodate the foregoing points, without surrendering Merleau-Ponty to the argumentative circularity of (C).

It might be thought that the *Phenomenology* can avoid strict commitment at the outset to the transcendental framework—rather in the way Kant describes his Copernican notion that objects conform to our mode of knowledge as a 'hypothesis' to be tested and proven—if its argument is read as a dialectic between, on the one side, the various species of objective thought, and on the other, Merleau-Ponty's transcendental metaphysics of pre-objectivity, which is concluded ultimately in the latter's favour.

There is however a double difficulty with this suggestion. First, if Merleau-Ponty's starting point consists of two equally weighted hypotheses, a justification is wanted for the parity of the initial weightings. Why should the Empiricist, who enjoys, as Merleau-Ponty himself concedes, *prima facie* agreement with common sense, accept it? Second, even if this difficulty is removed, we are no better off in understanding how the dialectic can be conducted to the satisfaction of the Empiricist,²² since as said earlier, each point claimed by Merleau-Ponty as an opening to the

²⁰ See VI 5–7, where Merleau-Ponty explicitly distinguishes the (non-transcendental) outlook that *accepts* the challenge of skepticism, from the deeper and correct (transcendental) approach that starts with 'the *problem of the world*', namely, 'to know precisely what the being of the world means' (VI 6). The latter may be said to 'reformulate the skeptical arguments outside of every ontological preconception' (VI 6).

²¹ Again showing Merleau-Ponty to be aware of the difficulty is his emphasis on the obliqueness of our access to pre-objective phenomena. Though perception 'as an encounter with natural things is at the foreground of our research' (VI 158), and we 'do indeed first have to fix our gaze on what is apparently *given* to us' (VI 159), our real task consists in responding to the 'interrogation' elicited by 'margins of presence' or 'references' in our experience (VI 159), for the universe of objective thought 'can tell us nothing' about pre-objectivity 'except indirectly, by its lacunae and by the aporias in which it throws us' (VI 157). The crucial point is that something philosophically weighty is required to redirect our attention from the 'foreground' of experience to its 'margins' and to allow the 'lacunae' and 'aporias' of objective thought to be apprehended.

²² The case is different, I am about to argue, with respect to the Intellectualist.

pre-objective will be regarded by the Empiricist as merely a cue for the revision of scientific theory; the theory of perception on its own will provide no resources for showing either inference, that of the Empiricist or that of Merleau-Ponty, to have the greater justification.

3.4

It thus seems fair to conclude that there is no logical gap between Merleau-Ponty's rejection of Empiricism and notion of attaining presuppositionless-ness, and his commitment to transcendental explanation. This squares with the fact that for the most part Merleau-Ponty's text does not differentiate between the tasks of criticizing Empiricism, Kant, etc., and of formulating his own transcendental theory: the relevant arguments are presented alongside one another rather than serially, so that the illegitimate hegemony of objective thought and the reality of pre-objective being are brought into view simultaneously.

If (C) is correct, and the *Phenomenology* is committed from its very first page to a transcendental framework which its discussion of perception presupposes, does the absence of an independent rebuttal of realist or naturalistic positions constitute a weakness in its argument? It does not, if the *context* of that argument is one in which it is already accepted that the proper form of philosophical explanation consists in the provision of transcendental conditions. In that case, the *Phenomenology* should be regarded as simply not addressed to the convinced naturalist or scientific realist: it is not intended to persuade anyone who is not already of a transcendental persuasion. Though this means that in one respect Merleau-Ponty is merely preaching to the converted, by no means does it render his argument pointless: the *Phenomenology* is targeted at those who *accept* the need for transcendental explanation but *identify* transcendental conditions in *objective* terms, and it is with respect to *these* positions that it is supposed to do its work. The reason why *non-transcendental* philosophy—naturalism, scientific realism—figures centrally onstage among Merleau-Ponty's targets is that he wishes to demonstrate *to* the transcendental philosophers of objective thought—Kant, Husserl, Sartre—that their own positions are unwittingly continuous with the non-transcendental positions that they reject, i.e. that their attempts to define their positions in opposition to transcendental realism and the philosophical prejudices of the natural attitude are only partially successful. Kant has not, Merleau-Ponty wishes to show, eradicated from his position all that the Copernican revolution was (or ought to have been) intended to overcome, while Husserl and Sartre have failed to extirpate elements of the natural attitude that their phenomenologies were intended to eliminate.

3.5

In section 4 I will cite a further passage bearing out these claims, but now I want to give a textual illustration which supports this point about the scope and assumptions of Merleau-Ponty's argument from a different angle.

In 'The Primacy of Perception and its Philosophical Consequences', published shortly after the *Phenomenology*, Merleau-Ponty offers a defence of the book's main theses, and in order to give an idea of how he conceives perception, summarizes his view of the problem set by 'an object which we perceive but one of whose sides we do not see' (*PrP* 13). Merleau-Ponty considers three philosophical analyses of this perceptual situation:

1. 'I *represent* to myself the sides of this lamp which are not seen.'
2. 'The unseen sides are *anticipated* by me.'
3. 'The unseen sides are simply *possible perceptions*.'

Merleau-Ponty rejects these analyses because they imply that the unseen sides are *not present* for me—they suggest either that the existence of the unseen sides is *merely probable* or that my relation to them is one of *mere belief*, i.e. mediated by a truth which has been grasped in the way that we grasp truths of geometry, in place of a direct relation to an object.

What analysis does Merleau-Ponty offer in their place? What he says is this:

The unseen side is present in its own way. It is in my vicinity . . . [When] I consider the whole setting [*l'entourage*; i.e. the relation to touch, etc.] of my perception, it reveals another modality which is neither the ideal and necessary being of geometry nor the simple sensory event. (*PrP* 14)

What is the nature of this 'other modality'? Here all that Merleau-Ponty does is to refer us to the further details of his discussion of perception and the body in the *Phenomenology*. But to say that the unseen side is 'present in its own way' obviously does not amount to an *analysis* at all, in the sense of those that he wishes to challenge. Merleau-Ponty does not, in fact, have a *rival* explanation of the cognitive achievement; rather he is *refusing the question*. Consequently the naturalist will regard Merleau-Ponty's argument as making no impact. But what this should really be taken to show is that Merleau-Ponty is inviting us to regard pre-objective perception as *not requiring* explanation or permitting analysis of the sort that Empiricists and Intellectualists are keen to offer—he is suggesting that we relocate it on the side of the philosophical *explanans*, and motivating this suggestion by indicating that the cost of not doing so is a reduction of the object's unseen side to a matter of mere belief. This makes full sense on the Transcendental Interpretation, since if perception is a ground-level transcendental condition, then it *could not* receive any explanation. The only alternative to this construal, I think, is to view Merleau-Ponty's argument here, and a great many others in the *Phenomenology*, as missing their target and as entirely beside the point. The Transcendental Interpretation may not give Merleau-Ponty an argument against the *naturalist*, but it does give him an *argument*, addressed to his transcendental predecessors.

3.6

I have argued that the *Phenomenology* cannot be divided into discrete steps providing a logically linear argument for the transcendental turn, but it can of course still be

considered how, *with* the transcendental framework presupposed, its argument may be intended to run. No doubt various reconstructions are possible, but the following is a rough sketch of the role played by the discussion of perception in the argument of the Introduction and Part One:

- (1) The nature of perception, correctly apprehended, invites us to consider it unanalyzable. [A phenomenological datum disclosed to phenomenological intuition in the course of the Introduction and Part One.]
- (2) Perception qua unanalyzable is suited to play a transcendental role, i.e. a candidate for the *explanans* in transcendental explanation. [Here the presupposed transcendental framework enters: Merleau-Ponty assumes the *need* for transcendental roles to be played, the task being to identify their occupants.]
- (3) Empiricism does not offer transcendental explanation.
- (4) Intellectualism offers an ostensibly transcendental theory, but of a kind which misrepresents the nature of perception and denies it a transcendental character. [Again, a phenomenological result of the Introduction and Part One.]
- (5) Empiricism and Intellectualism are led to affirm the analyzability of perception by their shared commitment to objective thought.
- (6) Transcendental explanation cannot take the form of—it is incompatible with—objective thought. [An argument to be examined in section 4.]
- (7) Transcendental explanation must lie in perception conceived pre-objectively.

On this reconstruction, the key connections hold between the notions of perception's unanalyzability, transcendental role, and pre-objectivity. What should next be emphasized is that it is not for Merleau-Ponty the whole story. The Introduction and Part One cannot be taken independently of the concluding chapters, in the context of which a much broader argument comes into view. Here is a rough reconstruction of that broader argument:

- (1) Human existence is characterized by specific forms of intersubjectivity, temporality, and freedom. [These specific forms are described in the concluding chapters, mainly on a negative basis, i.e. through criticism of naturalistic, Husserlian, Sartrean, etc., accounts of these topics.]
- (2) These specific forms of intersubjectivity, etc., cannot be grasped by means of objective thought. [As section 4 will elaborate, Merleau-Ponty tries to show, regarding each of the topics, that objective thought necessitates various alternatives each of which is unacceptable: for example, it implies that we have either no freedom or absolute unqualified freedom à la Sartre.]
- (3) The world in general must be interpreted in a way that explains how it is possible for human existence to be such that it cannot be grasped in objective thought. [In other words: it is not coherent to affirm that *human existence* has the specific character assigned to it, unless it is also affirmed that the *world* inhabited by human subjects has an appropriate correlative metaphysical character. Therefore:]

- (4) The world in general must be interpreted as being such that it cannot be grasped in objective thought.
- (5) The world in general must repose upon pre-objective being.

There are various ways in which this argument, and that of the Introduction and Part One, may be regarded as related. The concluding chapters, and the Introduction and Part One, can be regarded as two *parallel*, mutually supporting and illuminating, arguments for the same conclusion. Or the argument of the Introduction and Part One can be regarded as *embedded within* the argument of the concluding chapters—as elaborating its line (3). Or again, one could shift the whole centre of gravity to the concluding chapters and reduce the Introduction and Part One to a lengthy prolegomenon.

3.7

One central theme in the history of transcendental thought has been the search for internal connections between theoretical and practical philosophy, between metaphysics and the theory of value. It may appear, however, that nothing much by way of a practical theory or theory of value is contained in the *Phenomenology*, and that Merleau-Ponty forgoes the attunement of metaphysics to ethical issues trumpeted in *Being and Time* and *Being and Nothingness*. The appearance of being concerned exclusively with questions of theoretical philosophy no doubt encourages the Psychological Interpretation to treat the *Phenomenology* as first and foremost a study in the philosophy of mind.

There is no space here to consider the issue in full, but the following brief remarks are worth making to show that Merleau-Ponty does envisage connections of the metaphysics of the *Phenomenology* with practical and axiological issues.²³

In the first place, Merleau-Ponty's non-naturalistic idealist metaphysics set human existence at the centre of reality and dispose of the 'nihilist' threat posed in transcendental eyes by scientific naturalism. More specifically, Merleau-Ponty's ground-level reciprocal interweaving of self and world establishes a sense in which the human subject is fundamentally at home in the world, *bei sich*, in a way that contradicts the postlapsarian, arguably tragic or quasi-nihilistic portraits of the human condition painted by Heidegger and Sartre: for Merleau-Ponty, human being is not *constituted* by a metaphysical problem—of alienation from Being, or *manque d'être*—in the way that it is for Heidegger and Sartre.

Although Merleau-Ponty only hints at the connection (*PrP* 25), the *Phenomenology*'s conception of pre-objectivity provides for a primordially given unity of fact and value, akin to what is encountered in aesthetic consciousness. Because perceptual pre-objectivity comprises not bare sensory qualia but rather contains meaning,

²³ *SB* 223–34 shows Merleau-Ponty regarding the phenomenology of perception as an essential propaedeutic to moral theory.

scrutiny of the given does not drive us to suppose that value, or the source of what will come to be grasped reflectively in the form of values, is absent from it. The background value-permeation made possible by Merleau-Pontian pre-objectivity offers a foundation for moral realism (*PP* 456). This points away from the Psychological Interpretation: if Merleau-Ponty's conception of pre-objectivity incorporates or makes provision for value, realistically conceived, then it can hardly be *identified* with a richer conception of perceptual content.

Though Merleau-Ponty denies that determinate concrete practical implications can be *deduced* from his metaphysics, the specific forms of intersubjectivity and freedom described in the concluding chapters of the *Phenomenology* are not bereft of practical implications.²⁴ Merleau-Ponty stands opposed, as a moral particularist, to Kantian formalism and Sartrean voluntarism (*PP* 456), and his account of the shared intersubjective perspective contradicts Sartre's conflictual account of human relationships: before the for-itself's Look can begin the Sartreian dialectic of mutual objectification and counter-objectification, it is necessary that self and other co-perceive themselves as *sharing* a world in which each, as embodied, is not merely intersubjectively accessible to the other but also exists freely *in concert with* the other (*PP* 456); self and other may elect to negate one another, but it is not metaphysically necessary that they do so, contra Sartre (*PP* 448). Again, Merleau-Ponty claims to have exhibited the inescapability of responsibility in a way that objective thought fails to do: human freedom is protected against the Empiricist's reduction of my being to that of a mere 'thing', but it is not made to rest precariously, in Sartre's Intellectualist fashion, on my 'taking up' my natural and social facticity through a metaphysical intervention undertaken from outside being (*PP* 456).

The practical and axiological dividends of his transcendental metaphysics comprise a further important dimension of what I have called Merleau-Ponty's broader argument, and add extra weight to the Transcendental Interpretation.

3.8

The next point demanding emphasis concerns Merleau-Ponty's view of the nature and limits of philosophy. Merleau-Ponty's writings overflow with remarks about how ultimately philosophy can do no more than bring us to *see* how things are pre-objectively. Merleau-Ponty talks of phenomenology as 'restor[ing] the world of perception' (*PrP* 3), 'a method of getting closer to present and living reality' (*PrP* 25): '[t]rue philosophy consists in relearning to look at the world' (*PP*, xx); philosophy must 'conform itself with the vision we have in fact', 'adjust itself to those figured enigmas, the thing and the world' (*VI* 4); '[p]henomenology, as a disclosure of the world, rests on itself, or rather provides its own foundation' (*PP*, xx-xxi); it 'wishes to bring to expression' 'the things themselves, from the depths of their silence' (*VI* 4); 'philosophy has no other

²⁴ See *PP* 456, where Merleau-Ponty affirms the necessity of both 'willing freedom for all' and 'silence' on matters of ethical theory, and *PrP* 25-6.

function than to teach us to see [things] clearly once more, and . . . comes into being by destroying itself as separate philosophy' (*PP* 456).

The suggestion that philosophical cognition involves something extra-propositional cannot be missed. The further, key point is that this non-propositional something is regarded by Merleau-Ponty not merely as a necessary accompaniment or precondition of philosophical cognition, but as what philosophical cognition essentially *consists in*: having put 'the certainties of common sense and a natural attitude to things' 'out of play', suspending 'for a moment our recognition of them', we experience "wonder" in the face of the world', and 'from this break we can learn nothing but the unmotivated upsurge of the world'—we do no more than 'watch the forms of transcendence fly up like sparks from a fire' (*PP*, xiii–xiv). Such apprehension marks the *limit* of philosophy: '[a]ll that has to be done is to recognize these phenomena which are the ground of all our certainties'; belief in 'an absolute mind' or in 'a world in itself detached from us' is nothing more than 'a rationalization of this primordial faith' (*PP* 409). The rationality of our commonsense certainties 'is not a *problem*', for there is nothing 'behind it' for us to determine (*PP*, xx). We may call it a 'miracle' or 'mystery', but it is not one that leaves matters 'problematical': since 'we are ourselves' the 'network of relationships' which it establishes, 'nobody knows better than we do how this miracle is worked'; the mystery 'defines' the world and reason, so 'there can be no question of dispelling it by some "solution"' (*PP*, xx). To 'establish this wonder' is 'metaphysics itself' (*PrP* 11).²⁵

Merleau-Ponty thus belongs to a tradition which treats philosophical knowledge as consisting in the attainment of states of mind which consist in more than doxastic attitudes to philosophical propositions. These privileged cognitive states are in a limited sense ineffable: they can be registered linguistically, but their propositional expressions function only as indices. It is for this reason that Merleau-Ponty can claim, as we have seen him do, the discursive non-provability of his philosophical conclusions.²⁶ The view is also crucial for his idea that painting (Cézanne) may communicate the same content as phenomenological philosophy.²⁷

²⁵ See also *The Prose of the World*, ch. 4, where Merleau-Ponty makes it clear that he regards awakening mystery as the end-point of philosophy, and explains how this requires philosophical reflection and differs from returning to the natural attitude. It is also made clear (1974, 123–5) that the account of objective thought which Merleau-Ponty gives in the context of *perception* is the basis for (and presupposed by) his position on (all) other things—theoretical cognition in general, holding-true, is held to have the character that it does because and only because of the original, mystery-occluding taking-up of perception by objective thought. Merleau-Ponty's conception of mystery as the terminus of philosophical reflection appears to owe much to Gabriel Marcel.

²⁶ The comparison with Fichte and Schelling's grounding of philosophical thought in intellectual intuition is worth drawing; Merleau-Ponty himself interprets Schelling's conception of intellectual intuition in terms of pre-objective perception: 'l'"intuition intellectuelle", qui n'est pas une faculté occulte, mais la perception même avant qu'elle ait été réduite en idées' (1968, 107).

²⁷ See 'Eye and Mind', esp. 1960, 161, 166, 178–80, and 'Cézanne's Doubt', esp. 1945c, 14–16. In 'Metaphysics and the Novel', discussing philosophy's relation to literature, Merleau-Ponty describes philosophy as aiming to 'giv[e] voice to the experience of the world' and fulfil 'the task, not of explaining the world . . . but rather of formulating an experience of the world' (1945b, 27–8).

Merleau-Ponty's view of the intuitive nature of philosophical cognition evidently makes a crucial difference to how the Introduction and Part One should be understood: if the Psychological Interpretation were correct, then the phenomenologist's experience of perception's pre-objectivity would be mere data, mere *evidence* for some philosophical proposition, whereas Merleau-Ponty's claim, we have just seen, is the reverse—the experience *is* the terminus ad quem of philosophical activity. His statement that phenomenology 'restores the world of perception' means, therefore, not just that phenomenology shows the importance of perception for philosophy, but that its practice generates in the philosopher an actual awareness of perceptual experience which the philosopher grasps as *completing* the philosophical task.

4. Antinomy, Idealism, and Transcendental Ambiguity

4.1

Merleau-Ponty's extra-propositional conception of philosophical knowledge is bound up with his strategy of moving philosophy beyond the attempt to formulate discursive solutions for its perennial problems, by interpreting those problems as expressions of the inherent limitations of thought. Because this strategy is essential for grasping what exactly pre-objectivity amounts to for Merleau-Ponty, as well as providing powerful support for the Transcendental Interpretation, I will discuss it in some detail.

The strategy is best understood as a novel development of Kant's argument that transcendental idealism is uniquely capable of resolving philosophical problems which are otherwise insoluble. In the Antinomy of Pure Reason, Kant takes four topics in traditional metaphysics and in each case argues that a contradiction—e.g. both affirmation and denial that the world is finite in space and time—can be derived through valid arguments. The four antinomies exhibit a single general form of conflict in metaphysics, between dogmatic rationalism and skeptical empiricism. The rational response to this paradoxical situation, Kant argues, is to identify in each case some proposition which is presupposed by both sides but can be denied, and the denial of which eliminates the contradiction. The presuppositions of the four antinomies, according to Kant, revolve around reason's idea of the world as a determinate totality, but ultimately, he suggests, one unarticulated claim lies behind them all, namely that the objects of our knowledge are things in themselves, the defining claim of transcendental realism. On the basis that transcendental realism is sufficient as well as necessary to generate the antinomies, and that the only alternative to it is transcendental idealism, Kant claims to have provided an indirect proof of transcendental idealism.

In the following passage, Merleau-Ponty indicates the close relation between the *Phenomenology* and Kant's Antinomy of Pure Reason:

It is true that we arrive at contradictions when we describe the perceived world. And it is also true that if there were such a thing as a non-contradictory thought, it would exclude the whole of perception as simple appearance. But the question is precisely to know whether there is such a thing as logically coherent thought or thought in the pure state. This is the question Kant asked himself. . . . One of Kant's discoveries, whose consequences we have not yet fully grasped, is that all our experience of the world is throughout a tissue of concepts which lead to irreducible contradictions if we attempt to take them in an absolute sense or transfer them into pure being. (*PrP* 18)

The chief contradiction which Merleau-Ponty has in mind as arising when we describe the perceived world concerns the 'relation of the perceiving subject and the world', which, he says, 'involves, in principle, the contradiction of immanence and transcendence', i.e. conception of the objects of perception as both immanent to acts of perception and transcendent of them (*PrP* 13). In this sense, 'the perceived world is paradoxical', the 'perceived thing itself is paradoxical' (*PrP* 16).²⁸

Now there is no explicitly articulated argument in Merleau-Ponty which matches the formal rigour of Kant's Antinomy. Even in the concluding chapters of the *Phenomenology*, where Merleau-Ponty's antinomy-strategy is most clearly visible—as already said, opposing conceptualizations of freedom, temporality, intersubjectivity, and so forth are argued in the concluding chapters to exhaust the possibilities of objective thought, clearing the field for Merleau-Ponty's thesis of pre-objective being—there is no strict attempt at a *reductio ad absurdum*; Merleau-Ponty's emphasis is on the individually unsatisfactory character of the options presented by objective thought, not their jointly comprising strict contradictions. Nonetheless, the Kantian connection asserted by Merleau-Ponty is easily grasped:

(i) Just as the Antinomy shows that contradictions can be avoided only if we deny identity between the given empirical world and the world qua object of reason, so the *Phenomenology* shows that we must deny identity between the perceived world and the world as conceived in objective thought; Kant's argument establishes that the world given in sense experience is mere appearance, Merleau-Ponty's that it consists of pre-objective being. The conclusions drawn are opposed—pre-objective being lacks the objective-conceptual constitution of Kantian appearance—but the form and idealistic trajectory of the arguments are the same: both attempt to demonstrate a lack of fit between what is given and what is represented by our concepts, which is then argued to imply, first, that the objects of our experience lack the subject-independence

²⁸ See also *PP*, xiv: 'in order to see the world and grasp it as paradoxical, we must break with our familiar acceptance of it'; and *SB* 215.

which our concepts represent them as possessing, i.e. idealism, and second, a limitation and demotion of the power of thought, Kant's conclusion being that pure reason cannot grasp nature and Merleau-Ponty's that the perceived world eludes thought's objectification.

It is, therefore, as if Merleau-Ponty had applied to the Kantian faculty of *understanding* the strategy of argument which Kant applies to the faculty of reason, and subjected Kant's idealism to the sort of critique to which Kant subjects transcendental realism. The result is a new kind of transcendental idealism, as it were a transcendental idealism of the second degree, which denies not only that empirical reality can be grasped by concepts independent of intuition, but also that the perceived world owes its reality and cognizability to the Kantian conjunction of intuition with objectivity concepts. In this way Merleau-Ponty's concept of pre-objective being can be regarded as a further development of the Kantian concept of appearance.

(ii) In both cases, venerable philosophical problems are held to have been solved, or dissolved, through being referred back from the *objects* of experience, where earlier philosophers had supposed their solution must lie, to the *subject's* power of thought, which is made to take the blame for producing contradictions. In Kant's Antinomy the relevant problems are the 'cosmological' problems of traditional metaphysics, including the key opposition of freedom and nature. Merleau-Ponty—again as it were taking Kant a step further, and applying Kant's own strategy to Kant himself—argues that *all* of the problems of epistemology and metaphysics, including those that the *Critique of Pure Reason* claims to solve, disappear in the light of the discovery of pre-objective being. The problem of the external world is dissolved with the recognition that the perceived world is 'strange and paradoxical' (PP, xiii). The proper 'remedy to skepticism' in general (PrP 26) consists in accepting the perceived world as the foundation for 'all rationality', and allowing that it 'comprises relations and a type of organization' that the supervening 'world of ideas' can reflect only in the form of paradox, for which reason it is possible for knowledge to appear problematic: the paradoxes in question, which must be acknowledged as '*the very condition of consciousness*' comprise 'the justified contradictions of transcendental logic' (PrP 13, 16, 18–19). The given opposition (on which Sartre erects his metaphysics) of our mode of being, being-for-itself, to that of the objects of thought, being-in-itself, is to be treated as a function of a 'contact with being' that 'really is ambivalent' (VI 75). In the case of time, its 'ambiguity cannot be resolved, but it can be understood as ultimate, if we recapture the intuition of real time which preserves everything, and which is at the core of both proof and expression' (PP 394). The problem of other minds too, from the phenomenological standpoint, is relegated to a construct of objective thought: '[u]nder these conditions'—namely, of our pre-objective bodily and perceptual being, and correlative pre-personal selfhood—'the antinomies of objective thought vanish' (PP 351); in place of a discursively formulated philosophical

solution to the problem of other minds, 'we live through,' *nous vivons*, our inter-subjective situation (PP 359).²⁹

The opposition of realism and idealism is itself an antinomy of objective thought:³⁰ we 'leave behind the dilemma of realism and idealism' (PP 430), because 'the solution of all problems of transcendence' is available 'in the thickness of the pre-objective present' (PP 433). On the interpretation I am defending, this means of course that the realism/idealism opposition is subsumed in one sense, but not that Merleau-Ponty's position is in *all* senses beyond realism and idealism, since his own metaphysics recreate idealism at a higher level: the 'pre-objective present', appeal to which allows Merleau-Ponty's to position his transcendental idealism of pre-objectivity above the various oppositions of transcendental/empirical realism/idealism within objective thought, is itself an idealist notion.

Expressed in general terms, therefore, Merleau-Ponty's mode of solution to philosophical problems is to reduce them to representations produced by objective thought in its confrontation with pre-objective being; all that properly remains of them, after this structure has been grasped, is the recognition that there is in reality, i.e. in pre-objective being, an irresolvable 'ambiguity'. To the extent that the perceived world remains 'paradoxical' as opposed to merely displaying ambiguity, it is because and in so far as objective thought continues to cast a shadow; just as, according to Kant, dialectical illusion cannot be eliminated.

4.2

The claim that ambiguity resides in reality marks the point where Merleau-Ponty's antinomy-strategy extends beyond Kant's.³¹ Now it must be asked what it means to say that we should not seek to 'purge' objects 'of all ambiguity' (PP 11), that 'we must recognize the indeterminate as a positive phenomenon' (PP 6).

The natural place to start is with the famous Müller-Lyer lines, of which Merleau-Ponty says that they are neither of the same length nor of unequal length, and that they thereby show the presence in perception of a 'positive indeterminate' (PP 6, 11–12).

Certain construals of this claim can be ruled out immediately. Merleau-Ponty cannot simply be describing a psychological *explanandum*, nor can he have in mind a

²⁹ In fact Merleau-Ponty's view of the problem of solipsism is more complicated, for he holds that the 'difficulties inherent in considering the perception of other people did not all stem from objective thought, nor do they all disappear with the discovery of' the pre-objective (PP 356). This does not, however, affect the present point, since he does hold that the problem is *partly* solved through pre-objectivity and in any case uses its *total* insolubility from the standpoint of objective thought as an argument against the latter.

³⁰ See PP 317–34, 430–3, and SB 219.

³¹ Kant's account of empirical knowledge does not make room for positive indeterminacy in Merleau-Ponty's sense: indeterminacy enters for Kant only as a function of as-yet-incomplete determination, of the sort that, Kant says in the third *Critique*, occasions reflective judgement. Reflective judgement is constrained to realize systematic unity of *Erkenntnis*, and in principle no indeterminacies will remain at the end of the day.

hypothesized feature of subpersonal information processing, the idea perhaps that perceptual data goes through a point at which values of relative length are not yet assigned. Nor can Merleau-Ponty *merely* be asserting that questions of length come into play only when the categories constitutive of objectivity have been applied, for this on its own would just lead to the Kantian position that intuitions without concepts are ‘nothing to us’, which would contradict what Merleau-Ponty wishes to claim, namely, that indeterminacy or ambiguity is *actually perceived*.

What may be said in the first instance is that, at one level and in one sense, ‘ambiguity’ is a way of conceiving the phenomenon that Merleau-Ponty regards as evoked *from the angle of objective thought*—ambiguity implies the coexistence of different *determinate* meanings, whereas Merleau-Ponty’s own claim is that the phenomenon is *pre-* or *indeterminate*. Thus when Merleau-Ponty says that ambiguity is ‘real’ and yet ‘denied’ by objective thought, he is putting the point in the terms of the position he is attacking, in order to allege an antinomy in objective thought: his argument being that (1) the phenomenon is real, (2) objective thought can conceptualize it only as an instance in which two different and, in the Müller-Lyer case, *contradictory* determinate meanings are both realized, (3) objective thought is forced to either accept a contradiction or repudiate the phenomenon, and thereupon reveals its limitation.

The use just made of the concept of ambiguity is negative, and the next task must be to attempt to pinpoint the *positive* sense that it has for Merleau-Ponty—we need to say what it is about pre-objectivity in general that the specific ambiguity of the Müller-Lyer diagram brings to light. The key point, I take it, is the following. The Müller-Lyer diagram qua pre-objective phenomenon sponsors two specific determinate judgements and to that extent Merleau-Ponty must affirm that it is *related internally* to the relevant concepts of equality and inequality of length; yet he cannot want to say that the lines as they are pre-objectively *satisfy those concepts* in the way that empirically real objects can be judged to do so. The pre-objective phenomenon thus appears to be more than a blind intuition, but less than a conceptualized intuition; our awareness of it is *not unthinking*, but it lacks the form of *judgement*. If we now confront Merleau-Ponty with Kant’s Transcendental Deduction, and ask how he stands on the issue of the involvement of concepts in experience, it is obvious that there is only one thing that he can say: namely that the decompositional analysis of pre-objectivity into intuitional and conceptual components cannot be carried through.³² And this is exactly what Merleau-Ponty does say, in the many places where he explains how he wants to go beyond Kant, or ‘redefine’ the understanding and other Kantian notions:³³ the conditions of experience, he insists, cannot be analysed in terms of form and content, and the opposition of ‘perceptual life’ and ‘concept’ gains application only when

³² I wish to thank Christian Wenzel for comments which led me to correct my understanding of Merleau-Ponty on this issue.

³³ *PP* 49, 53, 126–7, 144, 220; see also the comments on rationality, *PP*, xix–xx and 430, and my n. 39 regarding the a priori.

analytical reflection has falsely dissected the intentional tissue of sense experience (PP 53, 126–7; VI 157–8). We encounter here, as Merleau-Ponty's metaphilosophy prescribes that we should, a final limit to what reflection can deliver: all we can say is that pre-objectivity possesses a 'formed-contentuality', a 'proto-conceptuality' or whatever, which grounds concept-application but is left behind in the transition to objective-conceptual form.

We can now understand the close, two-way connection drawn by Merleau-Ponty, which is at first glance puzzling, between indeterminacy and transcendental explanation.³⁴ His claim is not that there are really, from the transcendental standpoint, no determinate empirical facts, but that the experience of indeterminacy as afforded by the Müller-Lyer diagram exhibits, in a way that ordinary sense cognition tends to conceal, the distinctive quality, character, 'texture', or 'shape' of pre-objectivity, with which comes appreciation of its transcendental status.

This also makes clear why Merleau-Ponty's non-Intellectualist post-Kantian idealism does not relapse into empirical idealism. The ground of empirical reality for Merleau-Ponty is *pre-objective*. The empiricist's conception of experience, by contrast, is that of *unconceptualized intuition*, and this on Merleau-Ponty's account is a *derivation* from objective thought—what the classical empiricist calls 'experience' is a subjective residue scraped off the surface of a world conceived in accordance with objective thought.

4.3

Perceptual ambiguity or indeterminacy provides Merleau-Ponty with a general model for metaphysics. The real metaphysical ambiguity in things which generates the problems of philosophy and gives rise to opposing philosophical positions is, like the length of the Müller-Lyer lines, something that can only be intuited—in so far as we seek to take it up and articulate it in reflective judgement, we find ourselves in contradiction. This ultimately real, unanalyzable ambiguity is brute but not unintelligible, for we make it intelligible by recapturing the relevant pre-objective *Erlebnis*, by 'living' the ambiguity. Empiricism and Intellectualism may be understood on the analogy with the two judgements of the Müller-Lyer lines—Empiricism interprets the lines as having different lengths, and Intellectualism as having the same length. They are therefore not mere arbitrary mistakes: the phenomena of objective thought are 'not fictions, but firmly grounded' (PP 356). Merleau-Ponty's own metaphysical view corresponds to the conception of the diagram as exhibiting indeterminacy 'as a positive phenomenon', and his conception of the end-state of philosophical enquiry to that of perception of the lines as indeterminately both/neither equal and/nor unequal. Whatever is made of this far-reaching development of Kant's antinomy-argument, its importance for Merleau-Ponty is beyond doubt: without it, Merleau-Ponty has

³⁴ If classical psychology had grasped the transcendental status of the body, it would have been led to 'the world no longer conceived as a collection of determinate objects' (PP 92).

no philosophical position worth the name, and his claim to have moved philosophy beyond the solution of its traditional problems is hollow.

5. Phenomenology, Psychology, and Philosophy of Mind

5.1

It might be argued—against the Transcendental Interpretation, and contrary to what I have been assuming regarding the way that the philosophical options divide up—that the opposition of transcendental and naturalistic standpoints is not fixed in metaphilosophical stone, and that a better, more comprehensive and philosophically progressive interpretation of Merleau-Ponty will result from taking him to be aiming at a rapprochement or synthesis of transcendental philosophy with scientific psychology.³⁵ Merleau-Ponty, it may be said, turns to empirical psychology in the first instance to provide a robust anchor for phenomenological reflection, in response to the incoherence that has come to afflict the phenomenological project (*PP*, vii), but ultimately what the *Phenomenology* teaches is that the transcendentalist/naturalist dichotomy itself is yet another antinomy of objective thought which we can see our way beyond.

If the promised synthesis of transcendentalism and naturalism could be made plausible independently—no mean feat—then it would furnish the basis for a reconstructive interpretation of Merleau-Ponty, but it is not in Merleau-Ponty's own line of sight. Consider how Merleau-Ponty concludes the *Phenomenology*'s Introduction:

A philosophy becomes transcendental, that is to say radical, not by installing itself in absolute consciousness without mentioning the ways in which this is reached, but by considering itself as a problem; not by postulating a total making explicit of knowledge, but by recognizing as the fundamental philosophic problem this *presumption* on reason's part.

That is why we had to begin our examination of perception with psychological considerations. If we had not done so, we would not have understood the whole meaning of the transcendental problem, since we would not, starting from the natural attitude, have methodically followed the procedures which lead to it. We had to frequent the phenomenal field and become acquainted, through psychological descriptions, with the subject of phenomena, if we were to avoid placing ourselves from the start, as does reflexive philosophy, in a transcendental dimension assumed to be eternally given, missing the true problem of constitution. We could not begin, however, our psychological description without suggesting that once purged of all psychologism it can become a philosophical method. In order to revive perceptual experience buried under its own results, it would not have been enough to present descriptions of them which might possibly not have been understood, we had to establish by philosophical references and anticipations the point of view from which they might appear true. Thus we could begin neither without psychology nor with psychology alone . . . But now that the phenomenal field has been

³⁵ See Hoeller 1993, proposing an 'intertwining' of science and philosophy in Merleau-Ponty.

sufficiently circumscribed, let us enter this ambiguous domain, with the expectation that the psychologist's self-scrutiny will lead us, by way of a second-order reflection, to the phenomenon of the phenomenon, and decisively transform the phenomenal field into a transcendental one. (PP 63)

This passage does speak of a rapprochement of philosophy with psychological *description*—transcendental philosophy is to be corrected through attention to 'psychological considerations', and psychological description is to be purged of 'psychologism' by philosophical reflection—but what is envisaged is not the philosophical naturalist's fusion or joint partnership of philosophy with psychological *science*. The value of psychology in the sense of attention to 'psychological considerations' is as argued earlier to *reform transcendental philosophy* by freeing it from its Intellectualist misconception, and this involves no positive estimate of psychological science as an independent source of knowledge that philosophy ought to accommodate. Phenomenological conclusions do not depend for their truth, according to Merleau-Ponty, on the results of scientific psychology,³⁶ and nor does our knowledge of them.³⁷ Engagement with scientific psychology sharpens and refines our appreciation of psychological considerations, which in turn helps us to reach a position from which phenomenological truth can be grasped on the basis of an apodictic relation to the pre-objective, rendering transcendental reflection strictly independent of any application of the scientific method.

The large quantity of empirical psychology in the *Phenomenology* is therefore not a sign of its epistemological authority for philosophical purposes. On the contrary, what Merleau-Ponty wishes to see is a *transformation of psychology*—of that limited portion of it worth salvaging from Empiricism—into phenomenology.³⁸ Merleau-Ponty tries to play down the philosophically imperialist character of his view—he says that '[p]sychology as a science has nothing to fear from a return to the perceived world' (PrP 24) and talks of freeing Gestalt psychology from its scientific misconception—but in truth what his position demands is an assimilation of the recuperable part of psychological science to philosophy, going in the opposite direction from the union of psychology with philosophy advocated by philosophical naturalism.³⁹ As rapprochements

³⁶ See 1968a, 117–20, where Merleau-Ponty rejects Heidegger's dismissal of science but circumscribes its philosophical contribution negatively, and 132.

³⁷ De Waelhens, in an introductory essay to the 2nd edn of *Le Structure du comportement* authorized by Merleau-Ponty, describes this as 'an absolute misreading' (1963, xxvi–xxvii).

³⁸ See e.g. PrP 23–4, and also Merleau-Ponty's account of Husserl's relation to psychology in (1958). Kockelmans 1993 takes a view of the relation between philosophy and psychology in the *Phenomenology* similar to the one I describe.

³⁹ Merleau-Ponty's remarks on modal distinctions (in e.g. the passage quoted from PrP 14 concerning the 'modality' of the unseen sides of the lamp) and on the a priori/a posteriori distinction (e.g. PP 220–2) are not intended to collapse those distinctions in a way that would abolish the transcendental/empirical distinction, which would plainly make nonsense of much else that he says. A careful reading of PP 220–2 shows that Merleau-Ponty's aim is once again to free the transcendental a priori from its Intellectualist misconception, by stressing the immanence of the a priori in the object: the object has the a priori not as an 'external', in principle separable, superadded, formal component, but as its 'latent logic' (PP 221); the a priori transcendental elements 'in' objects present themselves 'spontaneously' in perception, not as subjectively originated external formal impositions. Connectedly Merleau-Ponty rejects the notion that philosophical reflection can through abstraction form concepts of transcendental a priori conditions which prescind from perception, in the way that Kant traces back the conceptual sources of the principles of possible experience to the

go, Merleau-Ponty's call for psychology to relate itself to phenomenological philosophy has more in common with Schelling and Hegel's positioning of *Naturphilosophie* within absolute idealism than it does with naturalism in the wake of Quine.⁴⁰

5.2

The Transcendental Interpretation restores to the *Phenomenology* the unity that Hyppolite queries and the Psychological Interpretation denies, and explains why for Merleau-Ponty there is no question of simply advancing, from the deficiencies of their extant forms, to improved versions of Empiricist and Intellectualist theories. If the Transcendental Interpretation is right, then it is incorrect to regard Merleau-Ponty's claims about perception as open to receiving support from branches of empirical psychology. The worry that Merleau-Ponty confuses epistemological with metaphysical priority is also resolved.

In light of the Transcendental Interpretation, it can be seen how the Psychological Interpretation arises and where precisely its mistake lies. The *Phenomenology* encompasses three sorts of claims: (i) negative critical claims about philosophical and psychological theories of perception, (ii) positive metaphysical claims about perception, and sandwiched between them, (iii) an extensive web of quasi-metaphorical descriptions of perceptual experience designed to elicit phenomenological intuition, which provide a two-way bridge between the work's critical and positive metaphysical claims—descriptions like that of the unseen sides of the lamp as being 'in my vicinity'. The Psychological Interpretation is generated by interpolating in this third intermediate domain positive theses about perceptual content of the sort found in the philosophy of mind—about its non-conceptual character and so forth.

One final comment on the Psychological Interpretation may be added. Theses in the philosophy of mind concern perception as commonsensically conceived. On Merleau-Ponty's account, however, this conception is contaminated by objective thought; it is what underlies Kant's conception of perception as empirical knowledge, *Erfahrung*. It follows that for Merleau-Ponty perception as ordinarily conceived

forms of judgement. Reflected in Merleau-Ponty's remarks on the *a priori* is therefore his 'revelatory' view of philosophical cognition discussed previously. Thus when Merleau-Ponty says that 'there is no longer any way of distinguishing a level of *a priori* truths and one of factual ones' (PP 221), he means that *there is* indeed an *a priori*, but that it cannot be determined at the level of consideration of what is necessary for thought of objectivity: grasping the *a priori* consists in the extraction of necessity from ('making explicit') concrete pre-objective experience; the *a priori* amounts to 'an absolute within the sphere of the relative' (PP 248). It is plain that the 'facts' with which Merleau-Ponty identifies the *a priori*, which incorporate a 'latent logic', are of no ordinary, naturalistic sort: 'the *fact* of sensory experience amounts to the assumption of a *form of existence*' (PP 221; emphasis added). Merleau-Ponty's transcendentalized conception of 'fact' is remote from any conception acceptable to philosophical naturalism, and his 'new definition of the *a priori*' (PP 220) brings him not a jot closer to philosophical naturalism than does Kant's definition.

⁴⁰ See n. 26 regarding Schelling. Merleau-Ponty approved Schelling's *Naturphilosophie* and claims Schelling as a precursor: see 1968a, 102ff. and 125ff., and 2003, 36–51. In view of Merleau-Ponty's denial that philosophical intuition can be recast in discursive systematic form and his assertion of the finality of paradox, affinities exist with German Romanticism as much as with German Idealism. This is rich territory which there is no space to explore here.

does not in the strictest sense exist. For this reason Merleau-Ponty talks of eschewing, at the beginning of philosophical enquiry, the very concept of perception: 'We exclude the term perception to the whole extent that it already implies a cutting up of what is lived into discontinuous acts' (VI 157–8). What does exist—and is shown to be the true, corrected referent of our ordinary ascriptions of perceptual states—is the pre-objective phenomenon of perception. The point, then, is not merely that Merleau-Ponty is not concerned with perception as the philosopher of mind attempts to theorize it, but that Merleau-Ponty considers that there *cannot be* a theory of perception, or positive theses about perceptual content, of the sort that the philosopher of mind attempts to provide. Merleau-Ponty's thesis of perception's pre-objectivity, and the philosophy of mind's thesis of the nonconceptuality of perceptual experience, do not therefore correspond: the latter concerns *content in the subject*, and the former the *world-as-perceived*, which in no sense lies in the subject. A non-metaphysical philosophy of psychology can be culled from the *Phenomenology*, as recent commentary has demonstrated, but it has only an oblique relation to the position that Merleau-Ponty is actually arguing for.⁴¹

References

- Baldwin, Thomas (1998), 'Merleau-Ponty', in Edward Craig (ed.), *The Routledge Encyclopedia of Philosophy*, vi (London: Routledge, 1998).
- Bermúdez, José Luis (2005), 'The Phenomenology of Bodily Awareness', in David Woodruff Smith and Amie L. Thomasson (eds), *Phenomenology and Philosophy of Mind* (Oxford: Oxford University Press).
- Bermúdez, José Luis, Anthony Marcel, and Naomi Eilan (eds) (1998), *The Body and the Self* (Cambridge, MA: MIT Press).
- Bickle, John, and Ralph Ellis (2005), 'Phenomenology and Cortical Microstimulation', in David Woodruff Smith and Amie L. Thomasson (eds), *Phenomenology and Philosophy of Mind* (Oxford: Oxford University Press).
- Dillon, M. C. (1998), *Merleau-Ponty's Ontology* (Bloomington, IN: Indiana University Press).
- Dreyfus, Hubert L. (1999), 'The Primacy of Phenomenology over Logical Analysis', *Philosophical Topics*, 27, 3–24.
- Dreyfus, Hubert L. (2005), 'Merleau-Ponty and Recent Cognitive Science', in Taylor Carman and Mark B. N. Hansen (eds), *The Cambridge Companion to Merleau-Ponty* (Cambridge: Cambridge University Press).
- Dreyfus, Hubert L., and Stuart E. Dreyfus (1999), 'The Challenge of Merleau-Ponty's Phenomenology of Embodiment for Cognitive Science', in Gail Weiss and Honi Fern Haber (eds), *Perspectives on Embodiment: The Intersections of Nature and Culture* (London: Routledge).

⁴¹ I wish to thank Mike Martin, Jerry Valberg, and others for their comments at the seminar at the University of London where an early draft of this chapter was presented, and the Arts and Humanities Research Board and the Philosophy Department of University College London for research leave that enabled its completion.

- Eilan, Naomi (1998), 'Consciousness and the Self', in José Luis Bermúdez, Anthony Marcel, and Naomi Eilan (eds), *The Body and the Self* (Cambridge, MA: MIT Press).
- Gallagher, Shaun (1998), 'Body Schema and Intentionality', in José Luis Bermúdez, Anthony Marcel, and Naomi Eilan (eds), *The Body and the Self* (Cambridge, MA: MIT Press).
- Gallagher, Shaun (2008), 'Cognitive Science', in Rosalyn Diprose and Jack Reynolds (eds), *Merleau-Ponty: Key Concepts* (Stocksfield: Acumen).
- Hammond, Michael, Jane Howarth, and Russell Keat (1991), *Understanding Phenomenology* (Oxford: Blackwell).
- Hoeller, Keith (1993), 'Phenomenology, Psychology, and Science, II', in Keith Hoeller (ed.), *Merleau-Ponty and Psychology* (Atlantic Highlands, NJ: Humanities Press).
- Hoeller, Keith (ed.) (1993), *Merleau-Ponty and Psychology* (Atlantic Highlands, NJ: Humanities Press).
- Kelly, Sean D. (2000), 'Grasping at Straws: Motor Intentionality and the Cognitive Science of Skillful Action', in Mark Wrathall and Jeff Malpas (eds), *Heidegger, Coping, and Cognitive Science: Essays in Honor of Hubert L. Dreyfus*, ii (Cambridge, MA: MIT Press, 2000).
- Kelly, Sean D. (2002), 'Merleau-Ponty on the Body', *Ratio*, 15, 376–91.
- Kockelmans, Joseph (1993), 'On the Function of Psychology in Merleau-Ponty's Early Works', in Keith Hoeller (ed.), *Merleau-Ponty and Psychology* (Atlantic Highlands, NJ: Humanities Press).
- Lacan, Jacques (1993), 'Merleau-Ponty: In Memoriam' (1961), in Keith Hoeller (ed.), *Merleau-Ponty and Psychology* (Atlantic Highlands, NJ: Humanities Press).
- Langer, Monika (1989), *Merleau-Ponty's 'Phenomenology of Perception': A Guide and Commentary* (Basingstoke: Macmillan).
- Merleau-Ponty, Maurice (1942), *The Structure of Behavior* [SB], tr. Alden Fisher (Boston: Beacon, 1963) [*La Structure du comportement* (Paris: Presses Universitaires de France)].
- Merleau-Ponty, Maurice (1945a), *Phenomenology of Perception* [PP], tr. Colin Smith (London: Routledge & Kegan Paul, 1962) [*Phénoménologie de la perception* (Paris: Gallimard)].
- Merleau-Ponty, Maurice (1945b), 'Metaphysics and the Novel', tr. Hubert L. Dreyfus and Patricia Allen Dreyfus, in *Sense and Non-Sense* (Evanston, IL: Northwestern University Press, 1964).
- Merleau-Ponty, Maurice (1945c), 'Cézanne's Doubt', in *Sense and Non-Sense*, tr. Hubert Dreyfus and Patricia Dreyfus (Evanston, IL: Northwestern University Press, 1964).
- Merleau-Ponty, Maurice (1947), 'The Primacy of Perception and its Philosophical Consequences' [PrP], tr. James M. Edie, in Merleau-Ponty, *The Primacy of Perception: And Other Essays on Phenomenological Psychology, the Philosophy of Art, History and Politics* (Evanston, IL: Northwestern University Press, 1964).
- Merleau-Ponty, Maurice (1958), 'Phenomenology and the Sciences of Man', tr. John Wild, in Merleau-Ponty, *The Primacy of Perception: And Other Essays on Phenomenological Psychology, the Philosophy of Art, History and Politics* (Evanston, IL: Northwestern University Press, 1964).
- Merleau-Ponty, Maurice (1960), 'Eye and Mind', tr. Carleton Dallery, in Merleau-Ponty, *The Primacy of Perception: And Other Essays on Phenomenological Psychology, the Philosophy of Art, History and Politics* (Evanston, IL: Northwestern University Press, 1964).
- Merleau-Ponty, Maurice (1964), *The Primacy of Perception: And Other Essays on Phenomenological Psychology, the Philosophy of Art, History and Politics* (Evanston, IL: Northwestern University Press).

- Merleau-Ponty, Maurice (1968a), *Résumés de cours: Collège de France 1952–1960* (Paris: Gallimard).
- Merleau-Ponty, Maurice (1968b), *The Visible and the Invisible: Followed by Working Notes [VI]*, tr. Alphonso Lingis (Evanston, IL: Northwestern University Press, 1968).
- Merleau-Ponty, Maurice (1974), *The Prose of the World* (c.1952), ed. Claude Lefort, tr. John O'Neill (London: Heinemann).
- Merleau-Ponty, Maurice (2003), *Nature: Course Notes from the Collège de France*, ed. Dominique Séglaard, tr. Robert Vallier (Evanston, IL: Northwestern University Press).
- Priest, Stephen (1998), *Merleau-Ponty* (London: Routledge).
- de Waelhens, Alphonse (1963), 'A Philosophy of the Ambiguous', tr. Alden Fisher, in Merleau-Ponty, *The Structure of Behavior* (Boston: Beacon).
- Wider, Kathleen (1997), *The Bodily Nature of Consciousness: Sartre and the Philosophy of Mind* (Ithaca, NY: Cornell University Press).

‘Hopelessly Strange’

Bernard Williams’s Portrait of Wittgenstein as a Transcendental Idealist

Stephen Mulhall

Recent discussions of the idea that there might be some illuminating relation between Wittgenstein’s later philosophy and that of Kant’s critical project have tended to take off from Bernard Williams’ highly influential paper ‘Wittgenstein and Idealism’ (a lecture he delivered in 1973 and published in 1974).¹ It is not often noted that this paper is, in effect, a critical review of Peter Hacker’s *Insight and Illusion*, published in 1972—a pathbreaking commentary on Wittgenstein’s writings, early and late, which explicitly organizes itself around the idea that Wittgenstein’s and Kant’s approaches to philosophy were importantly analogous.² And it is very rarely noted that Hacker was beaten to this particular insight by Stanley Cavell, who—in a famously devastating 1962 review of one of the first introductory books on Wittgenstein’s later philosophy (by David Pole)—explicitly related Wittgenstein’s idea of grammar to Kant’s idea of transcendental knowledge.³

Why should these matters of bibliographical priority be of any interest to us? First, they confirm that the idea of a comparison between Wittgenstein and Kant is not some relatively recent development in Wittgenstein scholarship, but rather something effectively coeval with it—as if unavoidably engendered by any serious study of Wittgenstein’s later work. Second, they raise the possibility that Williams’s understanding of Wittgenstein might be importantly inflected, to its detriment as well as its benefit, by Hacker’s sophisticated and highly influential but nevertheless contestable reading of Wittgenstein. Third, they might prompt us to recall that

¹ Reprinted in Williams 1981. The literature engendered by that article is immense: some of its more important reference points include Lear 1982, 1984; Bolton 1982; Sacks 2000, 6; Moore 2007.

² Hacker 1972.

³ Cavell 1962, reprinted in Cavell 1976, to which all page references are keyed.

both Hacker's and Cavell's ways of understanding this comparison are each rather more subversive than supportive of that developed by Williams. By the time he published a revised version of *Insight and Illusion* (in 1986), Peter Hacker had come to find this putative comparison so much more misleading than illuminating that he restructured the whole of his book so as to downplay it, and revised its subtitle accordingly. And in his 1962 review essay, Cavell no sooner identified the connection he saw between grammatical investigations and transcendental knowledge than he emphasized important differences between Wittgenstein's and Kant's ways of handling the very idea of such insights. Fourth, Hacker's and Cavell's qualms give us more than enough reason to contest Williams's (in fact rather half-hearted) attempts to argue that Wittgenstein's later work actually embodies a version of the first-person plural transcendental idealism whose basic structure it is the primary purpose of his paper to identify. And fifth, they also give us grounds for doubting the intelligibility of the very idea of transcendental knowledge as Williams develops it that are rather different from those he presents himself. This matters because, in my view, although he never exactly says as much, the conclusion of Williams's paper implies that his attempts to saddle the later Wittgenstein with a commitment to transcendental idealism are in the service of constructing an objection to those who might wish to take Wittgenstein's later conception of philosophy with any seriousness.⁴

1. Williams on the Early Wittgenstein

Williams begins his approach to Wittgenstein's later philosophy by specifying the sense in which he takes it that his early philosophy explicitly embodied a version of transcendental idealist thinking in its treatment of solipsism; for he intends to argue that, in his later work, Wittgenstein does not so much dispense with its three central features as transpose them into a first-person plural key. These three Tractarian ideas are:

[T]hat the limits of my language are the limits of my world; that there could be no way in which those limits could be staked out from both sides—rather [they] reveal themselves in the *fact that* certain things are nonsensical; and . . . that the 'me' and 'my' which occur in those remarks do not relate to an 'I' in the world, and hence we cannot conceive of it as a matter of empirical investigation . . . to determine why my world is this way rather than that way, why my language has some features rather than others, etc. Any sense in which such investigations were possible would not be a sense of 'my', or, indeed, perhaps, of 'language', in which the limits of my language were the limits of my world. (Williams 1981, 146)

⁴ A number of those well-acquainted with Williams and his writings have suggested to me that his essay was never intended to have such polemical significance. I find it hard to accept that suggestion, in part because of Williams's pretty much unwavering hostility to the work of Wittgenstein and his followers in his other writings; but even if the reader doesn't share my sense that this hostility is also present in this essay, I hope that my critical evaluation of it will retain some interest.

We can certainly see why this conjunction of ideas might be regarded as a species of transcendental idealism. The position they define is idealist insofar as it is an expression of solipsism, and transcendental insofar as the limits at issue are not empirical, and so not amenable to empirical investigation, and yet not what one might call transcendent either (not, that is, putative claims about a supra-empirical domain of entities beyond all possible knowledge).

What Williams is most concerned to note is that it is an implication of this conjunction of ideas that the insights it embodies are in fact inarticulable. 'We cannot in any straightforward sense say that there is, or that we can believe in, or accept, a metaphysical, transcendental self . . . ; for neither *what* it is, nor *that* it is, can be said, and attempts to talk about it or state its existence must certainly be nonsense' (Williams 1981, 146). Or, as Williams expresses it elsewhere, in the only sense in which we could meaningfully say that there is a transcendental self, that statement would be false (cf. Williams 1981, 153). And it is the reflexive problem that this creates for any such transcendental idealism that Williams will want to use in criticizing the first-person plural version of this position that he claims to detect in Wittgenstein's later work—what he describes in the conclusion to his paper as a new 'theory of meaning, [which] like the old, points in the direction of a transcendental idealism, and shares also the problem of our being driven to state it in forms which are required to be understood, if at all, in the wrong way' (Williams 1981, 163). After all, if any meaningful expression of such transcendental idealism is false, then any truthful expression of it is (not even false but) meaningless, necessarily incomprehensible, and so must be any theory of meaning that entails it. And this looks rather like a *reductio* of the theory.

As Williams notes, the author of the *Tractatus* is hardly unaware of this threatening feature of his 'theory': but on Williams's account, he embraces its apparently unpalatable implication, believing that philosophy can and must talk about this transcendental self 'in the only way in which by the end of the *Tractatus* we find that *philosophy* can talk about anything: that is to say, not with sense' (Williams 1981, 146). As Williams puts it later, 'the work itself tried, notoriously and professedly, to go beyond' the very limits of sense it had claimed to identify (Williams 1981, 161)—presumably to talk a kind of nonsense that is somehow illuminating, as opposed to mere gibberish (as Hacker has it in *Insight and Illusion*).

This is hardly an unfamiliar reading of the *Tractatus*. But it is, to say the least, a little surprising that Williams takes it for granted in this particular context, when focusing with such exclusive intensity on the solipsistic strand of this text's weave of thought. For in order to maintain this assumption, Williams must in effect reject the possibility that Wittgenstein actually means what he says in two sentences that Williams actually quotes from *Tractatus* 5.64:

[S]olipsism, when its implications are followed out strictly, coincides with pure realism. The self of solipsism shrinks to a point without extension, and there remains the reality co-ordinated with it. (Wittgenstein 1961, 5.64)

These sentences appear to point out that Wittgenstein's apparently solipsistic and hence idealist position turns out to be essentially indistinguishable from realism. And my purpose in citing them is not to suggest that Wittgenstein actually thinks of himself as a realist rather than a (solipsistic) idealist. It is rather to suggest that he seems to be pointing out in as unambiguous a manner as possible that the very idea of a distinction between idealism and realism (and so the very idea of a transcendental self by reference to which such a distinction might be drawn) turns out, when strictly followed through, to be simply and straightforwardly empty. It hardly seems to be the best way of advancing a case for transcendental solipsism to assert that the terms required for its mere articulation are devoid of sense.

A different reader of the *Tractatus* might have connected the idea of strictness invoked in 5.64 with its appearance in a related passage of the *Tractatus*—namely, in the claim advanced in 6.53 about what the strictly correct method in philosophy would be:

To say nothing except what can be said, i.e. the propositions of natural science, i.e. something that has nothing to do with philosophy: and then always, when someone else wished to say something metaphysical, to demonstrate to him that he had given no meaning to certain signs in his propositions. This method would be unsatisfying to the other—he would not have the feeling that we were teaching him philosophy—but it would be the only strictly correct method. (Wittgenstein 1961, 6.53)

How then, should we understand the presumably not strictly correct but more satisfying (and so perhaps more effective) method actually employed in the *Tractatus* itself? Perhaps as one in which the process of demonstrating to someone that they have failed to give meaning to some element in their utterance begins by treating their nonsensical utterance as if it were intelligible (thus giving him the feeling that genuinely metaphysical philosophy is being done), and then—by strictly following out its implications from within—brings the interlocutor to appreciate that he has failed to assign sense to some of the constituent terms of his utterance (say, 'idealist' or 'realist' or 'self'). If so, what Williams appears to regard as a devastating critique of the *Tractatus* would in fact be the *Tractatus's* own devastating critique of the metaphysical impulse in philosophy.

Perhaps it is unfair to expect even a philosopher of Williams's gifts to have taken seriously an interpretative possibility that has only been properly articulated in the last ten or fifteen years of *Tractatus* scholarship. Perhaps here we are confronting an unfortunate consequence of Williams's general (if not generally uncritical) reliance upon Peter Hacker's version of the more orthodox view of Wittgenstein's early attitude towards the unsayable. But one could surely expect Williams to have felt a little more acutely the pressure that (what he sees as) Wittgenstein's fundamental difficulty places upon his own work as an explicator of Tractarian ideas. The tension leaps out from this early conjunction of propositions within a single page of Williams's essay:

We cannot in any straightforward sense say that there is . . . a metaphysical, transcendental self; . . . attempts to talk about it or state its existence must certainly be nonsense . . .

[P]hilosophy can talk about it in the only way in which by the end of the *Tractatus*, we find that *philosophy* can talk about anything; that is to say, not with sense.

Whatever exactly we make of that, we can recover from the *Tractatus* discussion of the self and solipsism three ideas which will be particularly important points of reference in what follows: [and then we get the three points I quoted earlier]. (Williams 1981, 146)

In other words, within the span of a single sentence, Williams first registers his deep discomfort at the very idea of thinking that one can talk about the self and solipsism when any such way of talking must lack sense, and then tells us not only that we can recover three ideas concerning the self and solipsism from this nonsensical text, but also goes on to specify what they are. So are Wittgenstein's ideas about the self and solipsism lacking in sense or not? If they are nonsense, how can they be recovered, stated, and analysed? If they can be recovered, stated, and analysed, then what on earth does Williams mean by saying that transcendental solipsism of the Wittgensteinian variety is obviously and certainly unsayable? According to Williams's own account of the matter, any statement of these transcendental conditions that is meaningful must be false; and any accurate statement of them must be meaningless. So on which horn of this dilemma does he think that his own statement of these conditions falls? Neither seems exactly comfortable; but it is even more discomfiting to realize that Williams himself seems entirely oblivious to the need to choose between them—to the fact that anyone who wants to argue that Wittgenstein confronts a reflexive problem of this kind cannot do so without confronting the very same problem in the act of constructing that argument.

2. Williams on the Later Wittgenstein

So much for Williams's assumptions about Wittgenstein's early work; what are his reasons for claiming that the later work constitutes something other than an unequivocal extirpation of transcendental idealism—that its famous shift from 'I' to 'We' in fact '*takes place within the transcendental ideas themselves*' (Williams 1981, 147)? Well, the first point to note is that Williams doesn't exactly make any such claim. Indeed, he explicitly asserts that 'I am not going to claim anything as strong as that [Wittgenstein] held [a transcendental idealist view]' (Williams 1981, 153). But his reasons for restraint do not really grow from any doubt that the later Wittgenstein actually held such a view; they rather result from the assumption that any unqualified claim of that kind will be hard to substantiate—in part, because of the nature of the later Wittgenstein material (its inherent caginess as text), and in part because of the nature of the view (its inherent resistance to any formulation of it in propositions that are at once intelligible and true). Nevertheless, Williams does go this far, exegetically speaking:

I will offer some considerations which suggest that the influence of the sort of view I have sketched is to be felt in the later work, and that reference to it may help to explain some curious

and unsatisfactory features of that work. In particular, it may help us to understand the use that Wittgenstein makes of 'we'. (Williams 1981, 153)

The view Williams has just finished sketching in more abstract, non-exegetical terms, concerns an interpretation of the claim that 'the limits of our language mean the limits of our world' that is neither tautological nor empirical. The tautological reading would be: whoever 'we' refers to, it is going to be true that what we can speak of we can speak of, and what we cannot speak of we cannot speak of—true, but utterly trivial. The empirical reading would be: the 'we' refers to some particular group of language-users as opposed to another such group; so that, on a narrow construal of 'language' (taken to pick out one possible system of communication) together with a broad construal of 'world' (taken to pick out how in general the world appears), the claim would amount to the (Whorfian?) assertion that the way the world looks to different groups profoundly depends on what their language is like. Such an empirical claim would, on Williams's view, lead us to expect that the specific language used by a given group should provide some sort of an empirical explanation, however weak, of its way of looking at the world. But this is not the first-person plural idealism that Williams wishes to paint:

Under the idealist interpretation, it is not a question of our recognizing that we are one lot in the world among others, and . . . coming to understand and explain how *our* language conditions *our* view of the world, while that of others conditions theirs differently. Rather, what the world is for us is shown by the fact that we can make sense of some things and not of others: or rather—to lose the last remnants of an empirical and third-personal view—in the fact that some things and not others make sense. Any empirical discovery we could make about our view of the world . . . would itself be a fact which we were able to understand in terms of, and only in terms of, our view of the world; and anything which radically we could not understand because it lay outside the boundary of our language would not be something we could come to explain our non-understanding of—it could not become clear to us what was wrong with it, or with us. (Williams 1981, 152)

It is worth noting even at this early stage the speed with which the trivial reading of Wittgenstein's claim has disappeared from view. Williams seems to regard the triviality of the truths it would advance as sufficient reason for rejecting the idea that the later Wittgenstein's position might be accurately or exhaustively articulated in such terms—an assumption that flies in the face of Wittgenstein's own notorious claims that philosophy must leave everything as it is, that it simply puts everything before us, and that if one tried to advance theses in philosophy it would never be possible to question them because everyone would agree to them (Wittgenstein 1953, sections 124, 126, 128). These are certainly not either transparent or self-validating assertions; but however we might parse and justify them—a task to which Hacker devotes much time and effort in *Insight and Illusion*—they do not on the face of it place a trivial reading of anything the later Wittgenstein might claim entirely out of exegetical court.

But suppose we set aside this worry for the moment (we will return to it, at least obliquely, at the end of this chapter). Then we can put Williams's point this way. Anything about our view of the world that could be empirically explained could not constitute an element in the transcendental facts towards which the idealist is pointing, since the relevant features would fall within the world they are supposed to condition. So we cannot explain the world of our language in any such way; but we can clarify it, by self-consciously exercising it—'by moving around reflectively inside our view of things and sensing when one began to be near the edge by the increasing incomprehensibility of things regarded from whatever way-out point of view one had moved into' (Williams 1981, 153). And what one would thereby become conscious of is how we go on—which is a matter of how we think, speak, and intentionally conduct ourselves. So these transcendental facts are ultimately a matter of our interests, concerns, and activities (that is, not-further-explicable expressions of mind, and hence to be idealistically construed)—'although in the only sense in which we could meaningfully say that they determined everything that statement would be false' (Williams 1981, 153). In other words, even when it is transposed into first-person plural terms, such transcendental idealism cannot be both meaningfully and truthfully asserted.

If this is the kind of transcendental idealism with which Williams is concerned, how does he make good on his claim that its influence can be felt in Wittgenstein's later work, and that it might in particular explain the 'curious and unsatisfactory' use that Wittgenstein makes of 'we' in his grammatical investigations?

It has to be said that the exegetical argument has a bizarre construction. First, Williams simply asserts that Wittgenstein has little or no interest in 'language' understood in the narrow, Whorfian sense of that term, but is rather characteristically concerned with its extensive use, in which it embraces worldview rather than standing in explanatory contradistinction to it.

Hence his notoriously generous use of the expression 'language-game'; hence also in the converse direction, as it were, the tendency to use 'form of life' to refer to some quite modest linguistic practice . . . The narrower sense of 'language' seems not to be an important factor in any explanations Wittgenstein would want to consider for variations of world-view between human groups. (Williams 1981, 154)

This set of remarks is remarkably blithe in its freedom from any textual or logical tether. No exegetical evidence for either claim about Wittgenstein's use of these distinctive expressions is offered, despite the fact that much could be adduced against it (section 23 of the *Investigations* famously cites a multiplicity of language-games, none of which could with the wildest of imaginations be thought synonymous with 'our view of the world'). And even if it were the case that 'form of life' sometimes refers to some quite modest linguistic practice, why would that not show that quite modest, empirically distinguishable linguistic practices were always Wittgenstein's primary concern?

But Williams's broader claim is that Wittgenstein's lack of interest in language in the narrow sense hangs together with his utter lack of interest in the kinds of empirical

explanation in which such a use of the term might figure. More specifically, Williams thinks that Wittgenstein fundamentally rejects the evaluative comparability of different worldviews, and does so in a relativistic spirit—that is, in the belief that the question of whether something is empirically explicable or not is itself relative to a language (insofar as such explanatory practices are simply part of our language in the broader sense of the term). And this position, Williams has no difficulty in demonstrating, generates a large number of difficulties, of which the embrace of an aggregative solipsism seems only the most obvious. But what evidence is there that it is Wittgenstein’s position, as opposed to a position that some influential Wittgensteinians (for example, Peter Winch) have been taken as taking up? Williams cites absolutely none, taken up as he is with mapping the hypothetical confluence of natural currents of thought rather than attending to those actually running through particular texts of Wittgenstein (or, come to that, of Winch).

Williams then goes on to draw a peculiar conclusion from his mapping:

The difficulties we have now run into raise the question of whether Wittgenstein is really thinking at all in terms of actual groups of human beings whose activities we might want to understand and explain. I think the answer to that is basically ‘no’: . . . the alternative [ways of going on] are not the sort of socially actual alternatives, relativistically inaccessible or not, which we have been discussing, nor are they offered as possible objects of any kind of explanation. Rather, the business of considering them is part of finding our way around inside our own view, feeling our way out to the points at which we begin to lose our hold on it (or it, its hold on us) and things begin to be hopelessly strange to us. The imagined alternatives are not alternatives *to* us; they are alternatives *for* us, markers of how far we might go and still remain, within our world—a world leaving which would not mean that we saw something different, but just that we ceased to see. (Williams 1981, 159–60)

In other words, despite spending several complex pages on the matter, and despite charting passages of Wittgenstein’s writing in which the ‘we’ to which he is referring seems both to allow for intra-human linguistic difference and to suggest limits to our ability to evaluate such differences, it turns out that ‘relativism . . . is not really the issue’. Since it was only Williams who said that it was, the reader might be forgiven for wondering why he was led up this particular garden path; but since Williams has in so doing pointed out all the textual evidence that runs contrary to his imputation of relativism, the same reader might also be forgiven for wondering why he should think of it as a garden path (doubtless lined with primroses) at all. Why is it that ‘While the “we” of Wittgenstein’s remarks often looks like the “we” of our group as contrasted with others, that is basically misleading’ (Williams 1981, 160)? Why is it that ‘while much is said by Wittgenstein about the meanings we understand being related to our practice, and so forth, that we turns out only superficially and sometimes to be one we against others in the world’ (Williams 1981, 160)? What evidence does Williams have to offer that the way things so often seem to be is mere appearance?

To that question, there is precisely no exegetical answer to be found in the essay, now so close to its close. What we get instead is the claim that, because the later philosophy is committed to a particular assertion-conditions theory of meaning, and so to what Williams calls Wittgensteinian constructivism, this theory of meaning can avoid the threat of empirical idealism only by running into precisely the paradox that afflicts all the species of transcendental idealism that Williams has discussed. For ‘if our talk about number has been determined by our decisions, then one result of our decisions is that it must be nonsense to say that anything about a number has been determined by our decisions’ (Williams 1981, 163). In other words: if we could specify what did the determining, those determinants would have to be empirical facts about the history of our actual mathematical practice, and then empirical idealism would be unavoidable. The only way to avoid it is to adopt a perspective according to which any accurate specification of those conditions could not be said in anything other than the wrong way.

In short, Williams’s indictment of Wittgenstein as a first-person plural transcendental idealist depends upon the assumption that his later work embodies relativistic leanings, a hatred of natural science, and a constructivist theory of meaning. The charge is motivated not by any detailed analysis of Wittgenstein’s writings, but by the thought that cleaving to transcendental idealism would alone allow Wittgenstein to avoid difficulties generated by his relativism and his theory of meaning, as well as providing an apparent justification for (or perhaps just another expression of) his hatred of natural science. But since we have no good reason, certainly not any provided in Williams’s essay, for accepting that Wittgenstein is in fact committed to any of these three doctrines, the argument has precisely no force.

It is hard to avoid the conclusion that this essay is an exemplary instance of what can happen when someone knows before he begins to read Wittgenstein roughly what Wittgenstein will say; perhaps he knows this because of what Dummett or Hacker or Winch have already claimed to establish about Wittgenstein, perhaps because of other, less specific sources for the commonly available conception of ‘Wittgenstein’s later philosophy’ (common not only at the time of this essay’s composition, but also now). Whatever the precise aetiology of this presumptive knowledge, however, pure presumption it remains. Williams’s essay is essentially an indictment of Wittgenstein that is constructed and pronounced at second-hand—a rare example in Williams’s work of his allowing what everyone knows about an author to determine Williams’s own response to him.

3. (Dis)Inheriting Williams’s Wittgenstein: The Philosophical ‘We’

That said, it seems clear that the interpretative assumptions grounding Williams’s critique are broadly shared in the philosophical community, and that their further discussion might therefore raise matters of genuine interest for anyone wishing to

gain a deeper understanding of Wittgenstein’s real purposes and method. I have space to explore three of them: Wittgenstein’s supposed hostility to empirical science, his conception of grammatical limits, and the sense that his use of the first-person plural is somehow ‘curious and unsatisfactory’. Indeed, it can sometimes seem that most of the hostility and incomprehension with which Wittgenstein’s work is so often confronted in philosophical circles crystallizes around this third issue: call it the question of how his willingness to speak for himself (to speak personally rather than impersonally) gives him the right to speak for others.

First, however, I want to say a little more about Williams’s attribution to Wittgenstein of ‘a hatred of the cockiness of natural science that is hard to distinguish from a hatred of natural science as such’ (Williams 1981, 159). In the context of his essay, the kind of explanatory project that Williams suspects Wittgenstein is most concerned to quash is that which might hope to explain our way of looking at the world—in other words, attempts to account for the limits of our language by relating them to some or other feature of the way things actually are (our evolution, our environment, and so on). We have already seen that adopting transcendental idealism would block any such explanatory project; and Williams patently takes it that this feature of idealism would make it very appealing to Wittgenstein.

And yet, as Williams acknowledges, Wittgenstein in various places seems perfectly willing to allow for explanations of our having the concepts we do by reference to our interests, or the facts of human natural history, and so on. It is only at the very end of his essay that Williams implicitly acknowledges that Wittgenstein’s so-called hatred of natural science is explicitly registered in his philosophy (as opposed to his sporadic cultural reflections) only by his resistance to the thought that a particular conceptual system or grammar might be justified or validated by being shown to correspond to the way things are in reality. At this very late stage, Williams quotes a passage from *Zettel*, section 357: ‘We have a colour system as we have a number system. Do the systems reside in *our* nature or in the nature of things? How are we to put it?—*Not* in the nature of colours or numbers.’⁵

Williams’s gloss on this passage is tellingly confident in its diagnosis:

The diffidence about how to put it comes once more from a problem familiar in the *Tractatus*: how to put a supposed philosophical truth which, if it is uttered, must be taken to mean an empirical falsehood, or worse. For of course, if our talk about the numbers has been determined by our decisions, then one result of our decisions is that it must be nonsense to say that anything about a number has been determined by our decisions. (Williams 1981, 163)

⁵ I have corrected Williams’s misquotation of Wittgenstein’s actual remark. In the original, Wittgenstein ends by saying ‘*not* in the nature of numbers or colours’—a formulation which is far from synonymous with Williams’s version, in which Wittgenstein says ‘*not* in the nature of things’, as if assuming that numbers and colours are kinds of material entity.

Of course, if Williams had made any serious effort to establish that Wittgenstein was a constructivist theorist of meaning, or that he was a transcendental idealist for some other reason, then this flat assertion about the grounds of his diffidence might have some claim to serious attention. Since he has not, the best we can say for him is that at least he notes the diffidence. That is, he acknowledges that, in the relevant *Zettel* passage, Wittgenstein does not actually answer the question he poses; he rejects one possible answer to it, and otherwise registers bewilderment of a kind that suggests (as elsewhere in his later writings) a sense of latent obscurity or meaninglessness in the terms of the question.

If Wittgenstein *had* simply said that colour and number systems reside in our nature, in the minded domain of our interests, concerns, and activities, then an attribution of idealism would have some justification. But he specifically refrains from saying that; and he also avoids saying what Williams claims that he must be thinking if he is the peculiar kind of idealist he has laboured to characterize—namely that it would necessarily be false to say (what is nonetheless unsayably true) that the relevant conceptual system resides in our nature. He is, after all, perfectly willing to adopt exactly that stance about the apparently realist alternative—to declare it false that the relevant system resides in the nature of things. Neither way of putting it satisfies him: and since the form of the question allows for only those two ways of putting it, what he must be putting in question is the assumption that this question, so formulated, allows of an intelligible answer.

Why can we not say: these systems reside in the nature of what they allow us to represent? What sense of ‘residence’ is being rejected here? Well, since Wittgenstein is elsewhere happy to account for the structure of conceptual systems in terms of their utility for beings of given kinds in environments of given kinds, I would suggest that the kind of residence involved is likely to be that of correspondence. Wittgenstein’s thought is that grammatical structures do not stand in any kind of correspondence relation to reality, and so can neither correspond nor fail to correspond to the nature of things. In other words, they are not in the business of representing reality, but of providing means for the representation of reality; they articulate a field of meaning within which one can construct true or false descriptions of things, but they are not themselves descriptions of anything. This, I take it, is the burden of his discussion of ostensive definition and of rule-following, as interpreted by Peter Hacker: if definitions are not descriptions, and grammatical norms are akin to definition, then grammatical systems (understood as normative articulations of meaning or sense) cannot be a species of description. Our colour system is not the kind of thing that can either be justified or undermined by this kind of reference to reality; for in the terms articulated by which colour system would one describe that aspect of the nature of things to which a given colour system is claimed to correspond? The colour system under interrogation? Or some other one? Neither prospect seems inviting, if justification is supposed to consist of invoking something independent of that which is purportedly being justified.

So Wittgenstein's claim here is strictly limited—it concerns only the deployment of empirical knowledge in the service of a certain kind of philosophical project of conceptual validation, and so hardly amounts to the methodological registration of a hatred of natural science (cocky or otherwise). And its objection is to the intelligibility of any such project; so if the objection holds, it does not prevent us from doing something, but rather shows us that we have not yet given a clear meaning to our picture of 'conceptual systems being at home in reality', having reality as their correspondence address.

This is not a limitation, and so certainly cannot be thought of as a species of relativism (since it denies the intelligibility of a way of looking at the nature of things that was not grammatically articulable, and so deprives the relativist of the notion of non-relative understanding in relation to which alone our ways of making sense might appear relative). One might say that it is a limit; but if so, it is the kind that fences us off from nothing in particular. So it might be better not to invoke the notion of a limitation, or a limit, at all; and here we touch on the second of my themes: the idea of limits of sense, as reflected in Wittgenstein's conception of grammar.

It is in terms of this difference between a transcendental idealist commitment to undeniable but inarticulable limits and a deconstruction of the very idea of a limit that Cavell marks the uncanny, intimate difference between Kant's critical project and Wittgenstein's conception of philosophy in the essay I referred to earlier.

Where Kant speaks of 'transcendental illusion'—the illusion that we know what transcends the conditions of possible knowledge—Wittgenstein speaks of the illusions produced by our employing words in the absence of the (any) language game which provides their comprehensible employment . . .

If [t]his similarity to Kant is seen, the differences light up the nature of the problems Wittgenstein sets himself. For Wittgenstein it would be an illusion not only that we do know things-in-themselves, but equally an illusion that we do not (crudely, because the concept of 'knowing something as it really is' is being used without a clear sense, apart from its ordinary language game). So problems emerge [such] as: 'Why do we feel we cannot know something in a situation in which there is nothing it makes sense to say we do not know?' . . . (Cavell 1976, 65–6)

One might, of course, argue that there is a way of reading Kant such that the use to which he puts his concept of knowing 'things-in-themselves' moves him far closer to Wittgenstein's position than Cavell acknowledges. But then our understanding of transcendental idealism is very far from anything Williams officially invites us to associate with the term—as distant as a problem is from its solution, or rather an illusion from its overcoming.

I restrict myself here to what I called the official version of Williams's notion of transcendental idealism, because there are moments in his discussion of it which suggest that this notion is more internally unstable, or at least more capable of being read otherwise, than might at first appear. For example, Williams is eager to stress that Wittgenstein's supposedly non-empirical use of 'we' cannot even properly be

understood as restricting itself to human beings—as opposed, say, to nonhuman creatures (whether real or imaginary).

Nor is it just a question of a final relativization of ‘we’ to humanity. We cannot exclude the possibility of other language-using creatures whose picture of the world might be accessible to us. It must, once more, be an empirical question what degree of conceptual isolation is represented by what groups in the universe—groups *with* which we would be in the universe. If they are groups with which we are in the universe, and we can understand that fact . . . , then they also *belong* to ‘we’ . . . the plural descendant of that idealist *I* who also was not one item rather than another in the world. (Williams 1981, 160)

There is a certain asymmetry or nonreciprocity in this way of picturing the situation (a focus on what we can understand as opposed to what they can understand, and a lack of interest in, let us say, our mutual conversability) that is worth noting here, and to which I will return. But setting that aside, if Williams is right about this feature of the more universal use (what he thinks of as the only, or at least the obviously primary use) of Wittgenstein’s ‘we’, it is also worth noting a conclusion that follows pretty straightforwardly from it—one that Williams is himself quick to note, namely, that it would remain possible for us to explain what distinguishes the human view of the world from those of its nonhuman confreres (in terms of our differing evolutions, or environments, or whatever). And if that is possible should there be such nonhuman confreres, then it must be possible even if there are not:

Even if we, humanity, were the only lot in the world, a transcendental idealism of the first-person plural could not rule out in itself the possibility of an empirical or scientific understanding of why, as persons who have evolved in a particular way on a particular planet, we have the kind of world-picture we have—even though such an explanation would, once more, have to be within the limits of our language, in the only sense of ‘our’ in which they would mean the limits of our world. But if all that is possible, there is little left of the thought that those limits are *limits* at all: it might turn out with this sort of idealism, too, that ‘when its implications are followed out strictly, it coincides with pure realism’. (Williams 1981, 161)

This implicit citation of the *Tractatus* is also a reminder that all the evidence Williams marshals to convict Wittgenstein of transcendental idealism is in fact equally capable of showing that he transcends it. For if the course of reasoning Williams has just laid out is valid, it establishes that Wittgenstein’s purportedly universal use of ‘we’ neither excludes any intelligible form of empirical explanation of the human situation that we might care to construct, nor presupposes any sense of the limits of language which might coherently be called idealist (since the terms ‘idealism’ and ‘realism’ have no coherent, oppositional sense in this context). In short, this aspect of Wittgenstein’s later use of ‘we’ not only fails to justify the charges of hatred of natural science and idealism; it shows that neither charge, advanced on this basis, could conceivably be justified. At this point, Williams is able to resuscitate his indictment of Wittgenstein as an idealist in the final pages of his paper only by assuming, without any first-hand evidence, that Wittgenstein is committed to a constructivist theory of meaning. But

those of us unconvinced by any such assumption can safely leave the charge of idealism resting quietly in the grave it so naturally digs for itself.

Exactly the same barely repressed possibility of self-subversion also shadows Williams's most famous characterization of how the limits of sense reveal themselves in Wittgenstein's later writings—'in the fact that *certain* things are nonsensical' (Williams 1981, 146), something which becomes clear to us by 'moving around reflectively inside our view of things and sensing when one began to be near the edge by the increasing incomprehensibility of things' (Williams 1981, 153).

Take the general form of his claim first. In what sense, if any, is it a *fact* that certain things are nonsensical? Might they conceivably not have been nonsensical? If so, what sense would they have made? And if the relevant fact is that *certain* things are nonsensical, what things? If they were specific thoughts or propositions, then they surely can't be nonsensical; but if they truly are nonsensical, how can we talk of them as specific 'things' at all? Now look at the more specific form of that claim: if—in the relevant transcendental idealist sense of 'we'—we move around inside our view of things until we are near the edge, what exactly are we inside? If there is no outside for us to move, or not to move, around in, what sense attaches to the idea of being inside anything, and so of getting near the edge of that thing?

The terminology Williams employs to characterize the transcendental idealist's position persistently pictures the limits of sense as if they were limitations—as if they fenced us in, keeping us out of a domain beyond the domain whose limits they stake out, as if nonsense were a peculiar kind of sense, or as if there were something we cannot do here. But the whole thrust of the transcendental idealist enterprise as Williams defines it is to deny that anything lies beyond these limits, that they might be marked out from both sides; and the only consistent way of respecting that insight is to recognize that talk of inside as opposed to outside, or of limitations, or even of limits, is almost bound to be taken the wrong way.

The strictly correct solution to this problem is therefore not to continue using that language and those pictures, but rather to drop them altogether. The strictly correct embodiment of this insight is simply to point out to those who fail to acknowledge the difference between sense and nonsense when and how they have failed to assign a meaning to an element of their utterances; and to point out to those who try to acknowledge the difference between sense and nonsense by talking of limitations, limits, and moving around within those limits that they too have failed to assign any sense to their discourse about the limits of discourse. The strictly incorrect embodiment of this insight would be to begin by adopting these latently unintelligible ways of speaking, and then rendering patent their inevitably self-subverting tendencies. But neither approach would involve imagining; indeed could possibly involve imagining, that one might try to go beyond those limits. So why not assume that Wittgenstein was capable of seeing this?

So much for Wittgenstein's supposed hatred of natural science, and his supposed commitment to unsayably true grammatical propositions; what about his curious

and unsatisfactory use of 'we'? One assumption built into Williams's approach to this aspect of Wittgenstein's work is worth noting, and questioning, at the outset: namely, that there must be a single, fundamental, or determining pattern of use of the 'we' in Wittgenstein's later work. For Williams, either the 'we' refers to one human group as opposed to another (what one might call its empirical use), or it refers to the plural idealist 'we' to which there is no contrasting subject, whether individual or collective. But why assume that Wittgenstein's 'we' has a single, fundamental pattern of use throughout his later writings, or indeed within the *Investigations*? The sheer variety of phenomena referred to by such concepts as 'language-game', 'practice', and 'form of life' in that text would rather suggest a correspondingly various range of reference for his 'we'—from picking out the participants in highly localized and specific language-games, whether real or imaginary, from those who do not so participate (e.g. those who pray, or those whose concept of pain yokes together internal pains and those engendered by external injuries, as opposed to sharply distinguishing them; cf. *Zettel*, 380), to picking out patterns of linguistic activity that appear to be constitutive of anything we could recognize as part of a human form of life at all (e.g. those capable of describing an object or state of affairs, or of giving expression to their pain).

In the former kinds of context, Wittgenstein's purpose is patently to point out the sheer specificity of the relevant language-game, and so its difference from other, apparently cognate language-games (since he takes the philosophical problems in these locations to arise from the assumption that these games are essentially similar)—the 'I'll teach you differences' aspect of Wittgenstein's method. In the latter kind of context, his purpose is to show us that some aspects of our life with language are far less easy to imagine otherwise than might at first appear (as with our conception of what is to count as thinking or reasoning)—call this the 'the river-bed is not the river' aspect of his method. I know of no obvious reason why it should be illegitimate for Wittgenstein to deploy the term 'we' right across the spectrum of such examples. And if so, then careful attention to the specific context of the pronoun's employment will be vital in understanding its specific philosophical significance.

With this qualification in place, however, it has to be acknowledged that there is an aspect of Williams's resonant way of imagining the later philosophy's first-person plural investigation of the limits of sense—as a matter of feeling one's way around—that does cast important light on Wittgenstein's use of 'we'; for it conjures an impression of grammatical investigation as essentially experimental, a matter of trial and error, a procedure which leaves the exact point at which intelligibility runs out as an open question, not to be settled in advance of testing specific claims to have located those points.

This vision of the mode or form of Wittgenstein's later thinking patently conflicts with other ways of imagining grammatical investigations—ways which think of the grammar under investigation as having always already set limits to which our investigations can at most and at best recall us, as if returning us to the details of a book of rules, submission to which is a condition of responsibility as a member of the linguistic community.

We might think of this as tending to transform the personal force of Wittgenstein's 'we' into the essentially impersonal 'one'; more precisely, one might say that the limits of this 'we' are imagined as having been determined prior to any use of it in grammatical remarks. It is the transitional moment captured in Williams's remark, purportedly uttered from within the developing perspective of transcendental idealism, that 'what the world is for us is shown by the fact that we can make sense of some things and not of others: or rather—to lose the last remnants of an empirical and third-personal view—in the fact that some things and not others make sense' (Williams 1981, 152).

It is not hard to see how such an impersonal use of 'we' (the kind of philosophical use that is certainly to be found in *Insight and Illusion*, although not only there) might appear both curious and unsatisfactory to a commentator—to the point of forcing upon him the conclusion that its apparently ungrounded claim to authority could only be imagined to be justifiable if it is thought of as registering the existence of a plural subjectivity whose limits are fixed independently of experience and in such a way as to include all possible language-using creatures. Williams's construction of a first-person plural form of transcendental idealism would then amount to an attempt to envision the kind of linguistic subject who would be constituted by subjection to such a rulebook. His retention within that vision of an element of exploration, experimentation, and imagination undeniably runs counter to that vision, and threatens its internal coherence. But it shows an admirable awareness on his part of an aspect of Wittgenstein's use of 'we' that is not only absent from, but importantly repressed by, those who envisage grammar as an impersonal determination of sense.

This aspect is captured in Cavell's lifelong characterization of this philosophical 'we' as entering a claim to community. In his early essay, he puts the point this way:

Someone may wish to object: 'But such claims as "We say . . .", "We are not going to call . . .", and so forth, are not merely claims about what *I* say and mean and do, but about what *others* say and mean and do as well. And how can I speak for others on the basis of knowledge of myself?' The question is: Why are some claims about myself expressed in the form 'We . . .' ? . . . Then suppose it is asked: 'But how do I know others speak as I do?' About some things I know they do not; I have some knowledge of my idiosyncrasy. But if the question means 'How do I know at all that others speak as I do?' then the answer is, I do not. I may find out that the most common concept is not used by us in the same way. And one of Wittgenstein's questions is: What would it be like to find this out? . . .

If the little I have said makes plausible the idea that the question 'How do we know what we say (intended to say, wish to say)?' is one aspect of the general question 'What is the nature of self-knowledge?' then we will realize that Wittgenstein has not first 'accepted' or 'adopted' a method and then accepted its results, for the nature of self-knowledge—and therewith the nature of the self—is one of the great subjects of the *Investigations* as a whole. (Cavell 1976, 67–8)

On Cavell's reading, then—surely not unfamiliar to Williams, who dedicated his book on Descartes to Cavell—the first-person plural form of grammatical remarks is not meant to presuppose agreement on the part of some given range of interlocutors, but rather to test the extent of such agreement by articulating the philosopher's individual judgements about the points at which sense runs out, and inviting the other or others

to disclose or discover whether their individual judgements match, and so whether a community of speakers exists (in this specifiable respect). And it is precisely this aspect of his method that helps to give the whole field of interpersonal relations so prominent a place in his philosophical writings—as if his interest in that field is in part to be understood as an acknowledgement of what is controversial in that method, and an expression of a desire to reflect philosophically upon it (to make the sheer possibility of doing philosophy his way a central topic within its work—the medium as the question rather than the message).

Elsewhere, Cavell explicitly aligns this reading of the Wittgensteinian ‘we’ with Kant’s distinction between reflective and determinant judgement, as if grammatical remarks were akin to aesthetic judgements; and he thereby makes patent the internal relation between this conception of grammatical investigation and the imagination. The point is not that grammar is made up as we go along—Humpty Dumpty’s theory of meaning, as it were. It is rather that sense can be made wherever two competent speakers find that they can make it, which means finding that they can make sense of each other, and can go on together from their findings. On this picture of the limits of sense, they are not limits in any normal sense of that term—they fail to fence us in—because they are not so much fixed as to be fixed.

Cavell’s way of interpreting Wittgenstein’s ‘we’ thus devolves an ineliminable moment of responsibility upon every individual speaker who undertakes to employ it—a willingness to take responsibility for one’s ways of going on, and refusing to go on, with words that cannot ultimately be sloughed off on anyone or anything outside the interlocutory relationship within which that ‘we’ is employed. It thereby returns us to an implication of Wittgenstein’s repeated disavowal of any body of distinctively philosophical expertise—his conception of philosophical authority as finding its ground in the capacities of any competent speaker. For if that is the case, then a philosopher’s judgement of what makes sense has no less, but also no more, authority than that of any speaker; and so his interlocutor has just as much authority to contest or revise that judgement. To put matters the other way around: the philosopher’s claim to community can only be shown to be justified insofar as his interlocutor acknowledges it—recognizes himself in the other’s expression of his own sense that sense is running out, here and now. This is the very reverse of the idea that certain things just make sense, and certain other things don’t; and its conception of linguistic community is one which demands not the subordination of individual judgement but its full and free expression. What turns out to be ours is both, and can only be, yours and mine. But who exactly ‘we’ may be is itself at issue in every philosophical exchange; it is neither transcendently nor empirically predetermined, but rather to be worked out in the work of philosophy.⁶

⁶ This chapter was initially prepared for a one-day conference on ‘Wittgenstein and Transcendental Idealism’, held at the University of London in Jan. 2007, as part of a larger research programme on ‘Transcendental Philosophy and Naturalism’. My thanks to Mark Sacks, the director of this larger project, for his invitation to participate in this phase of its activities.

References

- Bolton, Derek (1982), 'Life-Form and Idealism', in Godfrey Vesey (ed.), *Idealism Past and Present* (Cambridge: Cambridge University Press).
- Cavell, Stanley (1962), 'On the Availability of Wittgenstein's Later Philosophy', *Philosophical Review*, 71, 67–93; repr. in Cavell (1976).
- Cavell, Stanley (1976), *Must We Mean What We Say?* (Cambridge: Cambridge University Press).
- Hacker, P. M. S. (1972), *Insight and Illusion* (Oxford: Oxford University Press). Revised edition published in 1986.
- Lear, Jonathan (1982), 'Leaving the World Alone', *Journal of Philosophy*, 79, 382–403.
- Lear, Jonathan (1984), 'The Disappearing "We"', *Proceedings of the Aristotelian Society Supplementary Volume*, 58, 219–42.
- Moore, A. W. (2007), 'Wittgenstein and Transcendental Idealism', in Guy Kahane, Edward Kanterian, and Oskari Kuusela (eds), *Wittgenstein and his Interpreters* (Oxford: Blackwell).
- Sacks, Mark (2000), *Objectivity and Insight* (Oxford: Oxford University Press).
- Williams, Bernard (1981), 'Wittgenstein and Idealism', in *Moral Luck: Philosophical Papers 1973–1980* (Cambridge: Cambridge University Press).
- Wittgenstein, Ludwig (1953), *Philosophical Investigations*, tr. G. E. M. Anscombe (Oxford: Blackwell).
- Wittgenstein, Ludwig (1961), *Tractatus Logico-Philosophicus*, tr. David Pears and Brian McGuinness (London: Routledge & Kegan Paul).
- Wittgenstein, Ludwig (1981), *Zettel*, 2nd edn, tr. G. E. M. Anscombe (Oxford: Blackwell).

Stoic Transcendentalism and the Doctrine of *Oikeiosis*

Wayne M. Martin

Transcendental inquiry is a philosophical strategy that has long been associated with the work and legacy of Immanuel Kant. Both Kant and his early sympathizers did much to encourage the view that their transcendental approach represented a radical departure from traditional philosophical aims and methods. It is certainly not my purpose here to challenge Kantian claims to originality. Nonetheless, I believe that there is an ancient precedent for Kant's transcendental approach: we find among the Stoics a strategy of investigation and argument that anticipates crucial elements of what in recent years has been dubbed 'transcendental argument' or 'transcendental proof'. By investigating this ancient history we can better appreciate both the resources of transcendental investigation itself and the distinctive contribution that Kant made to an ancient philosophical tradition.

Stoic influences on Kant are widely recognized, but are generally associated with a cluster of narrowly ethical doctrines regarding, inter alia, the motive of duty, the stringency of moral demands, and the cosmopolitan ideal. For present purposes, however, the pertinent doctrines are those that centre around the distinctive Stoic notion of *oikeiosis*. There is no clear modern analogue for this ancient notion, which no longer enjoys broad currency beyond classical scholarship, despite its centrality for Stoic doctrine and its influence on a number of subsequent traditions. Accordingly, our initial task in what follows must be to determine just what *oikeiosis* is. The first stage of my discussion is largely devoted to this question. In section 1 I survey the history of attempts to translate *oikeiosis*, and I review some standard philological information concerning its origin and use. In the second and third sections I consider its inferential role in two prominent Stoic arguments. In the fourth section I consider whether and in what sense *oikeiosis* is or involves a form of self-consciousness, and I propose an interpretation of *oikeiosis* as a form of normative orientation rooted in what I shall call

‘ontological self-consciousness’. With these results in hand I propose an interpretation (section 5) of the Stoic claims that *oikeiosis* serves as a (transcendental) condition on the possibility of intentionally determinate perception and desire. I conclude (section 6) by entertaining some objections, and by considering the relevance of these Stoic arguments for our understanding of the usual contrast between empirical or naturalistic claims on the one hand and transcendental claims on the other.

One word of warning is in order before turning to the matters at hand. Although Stoicism was once celebrated for its broad and systematic approach to philosophy, it is now generally remembered more narrowly as an ethical doctrine. Both in broader traffic in common usage and more narrowly within philosophical discussion, the Stoics are chiefly associated with their provocative ethical teachings: that virtue is the only genuine good and suffices to ensure happiness; that accordingly nothing can harm a good man, who will be ‘happy even on the rack’; that pleasure is not a genuine good nor pain a genuine evil; that life is to be lived ‘in accordance with nature’. At various junctures my discussion will touch quite closely on Stoic ethical teachings. But I wish to emphasize that my chief concern in what follows is not to assess or defend Stoic ethics, about which I here seek to remain studiously neutral. My aim, rather, is to extract from these Stoic discussions an underlying theory of the nature and function of self-consciousness, and to consider the role played by proto-transcendental arguments in advancing and defending that theory.¹

1. *Oikeiosis*: A Philological Primer

I start with the basics, with apologies to those for whom this part is old hat. What is *oikeiosis*? As it happens, almost everything about its interpretation is controversial. Some say it was the basic notion in Stoic ethics; others deny this.² Some say it was an idea original to the Stoics; others insist it was derivative.³ To some it marks a decisive step away from a persistent moral failing in ancient ethics; others argue that it was a dangerous and absurd doctrine.⁴ To get a first perspective on both the concept and some of the attendant difficulties of interpretation, it may be worth beginning with a

¹ A note about critical apparatus. Except where otherwise indicated, citations of classical sources refer to the editions of the Loeb Classical Library. I leave the Greek word *oikeiosis* untranslated, though I survey below some of the history of attempts to translate it, first into Latin and more recently into English. In order to avoid undue cluttering of the pages, I have not followed the usual practice of using quotation marks to distinguish uses from mentions of this concept; I hope and expect that the difference will in each case be clear from the context. Finally, I have throughout quoted from standard English translations of classical sources, both out of a realistic sense of my own qualifications to improve upon the experts, and because the modern rendering of ancient concepts is part of what concerns me here. The disadvantage of this policy is a loss of consistency and uniformity in the translation of certain technical terminology. I have tried to compensate for this loss by including key terms from the original texts in parentheses.

² For the affirmative claim see Pohlenz 1940 and Pembroke 1971; for the denial see Striker 1983, 165.

³ Von Armin 1926 argues that the Stoics borrowed the notion from the Peripatetic school; the Stoic claim to originality is vigorously defended in Pohlenz 1940. For a more qualified defense see Brink 1956.

⁴ For a sampling of this debate see Pangle 1998 and Padgen 2000.

selective survey of the history of translations. When Cicero set out to translate Greek philosophy into Latin he rendered the Greek term with a pair of Latin words: *conciliatio et commendatio*.⁵ Centuries later Grotius preferred to leave the Greek term untranslated, but explained it in Latin as *appetites societatis*—a desire for society.⁶ In the nineteenth century William Whewell translated *oikeiosis* as *the domestic instinct*, bringing out the connection to the Greek root *oikos*: a house or dwelling.⁷ Among recent classical scholars there has been an array of proposed translations: Long uses *appropriation*; Annas prefers *familiarization*; Schofield uses *affinity*.⁸ Pembroke leaves the term untranslated but glosses the term as *well-disposedness*.⁹ One standard modern translation of Greek sources deploys a range of terms: *affection*, *endearment*, and *being-near-and-dear*.¹⁰ In his translation of Plutarch, Cherniss renders *oikeiosis* as *congeniality*.¹¹ Both the lack of any settled translation and the diversity of these proposals provides us with a first hint that the term may not be easy to appropriate directly into an idiom shaped by modern assumptions—although even this claim has been challenged in the literature.¹²

In the face of the difficulties occasioned by this unfamiliarity and lack of consensus, my main approach here will be to resort to what is nowadays known as inferentialist semantics. We can fix the content of a concept, according to this approach, by uncovering its inferential role. What considerations are used in justifying the application of the concept? What does the invocation of *oikeiosis* entail? What role does it play in Stoic arguments and proofs? If we can identify the inferential patterns in which the Stoics themselves deployed the notion then we shall be well on our way towards understanding it for ourselves. Fortunately for this inferentialist approach, the Stoics were notoriously disputatious philosophers, and moreover were constantly defending their philosophical views in the face of a sustained barrage of criticism from rival schools. So there is no shortage of inferences to examine. In the sections that follow I consider two such inferential contexts in which the notion of *oikeiosis* occurs, the first pertaining to Stoic cosmopolitanism in ethics, the second to the Stoic theory of motivation and agency. Before turning to these disputes, however, we will do well to supplement our inferentialism with a somewhat more conventional philological approach.

According to the *Lexicon*, *oikeiosis* is to be defined as ‘a taking as one’s own, appropriation’.¹³ It derives from the root *oikos*, meaning house or dwelling—the same Greek root as in the more familiar modern word, ‘economics’ (literally: the rule of the

⁵ Cicero, *De Finibus* 3.16. Rackham translates this phrase as ‘attachment and affection’. More literally one might say ‘a bringing together and recommendation’. Cicero’s rendering has recently been echoed by Irwin, who translates *oikeiosis* as ‘conciliation’ (Irwin 2003, 252).

⁶ Grotius 1625, 6.

⁷ Whewell 1853, p. xxiv.

⁸ Long and Sedley 1987, i. 351; Annas 1992, 56–7; Schofield 2003, 243.

⁹ Pembroke 1971, 116.

¹⁰ Diogenes Laertius, *Lives of Eminent Philosophers*, tr. R. D. Hicks, 7.85.

¹¹ Plutarch, *On Stoic Self-Contradictions*, tr. H. Cherniss, *Moralia* 1038b.

¹² See Pembroke 1971, 114: ‘*Oikeiosis* does not need bringing up to date’.

¹³ Liddel and Scott, *An Intermediate Greek-English Lexicon* (Oxford, 1889, 1992), 545.

household). S. G. Pembroke provides a useful set of notes on the grammar and history of the term:

The verb *oikeioun*, which is transitive, turns up in various forms in the fifth and fourth centuries B.C. meaning to appropriate goods or, when applied to people, to win them over—the object of official diplomacy as well as of private intrigue . . . [T]he noun *oikeiosis* is used in this sense by Thucydides . . . *Oikeios*, to go back to the adjective, is regularly contrasted with *allotrios*, what belongs to someone else or is in wider sense alien to oneself[.]¹⁴

Pembroke's final comment here will be particularly useful to us in what follows, and already amounts to a first data point in our hunt for inferential patterns. For while the term *oikeiosis* may nowadays be unfamiliar, its traditional opposite is not. *Oikeiosis* is the opposite of *allotriosis*, the Greek term for alienation. As in modern English, these terms have an original economic sense: one alienates a piece of property by selling it, and one appropriates it when one makes it one's own. But as we shall see, these opposed economic terms come to have an extended application in both ethics and psychology.

The early history of the use of *oikeiosis* in philosophy is uncertain and disputed, but most scholars credit its introduction specifically to the Stoics. The word is absent from the extant writings of Plato and Aristotle, but it can be traced at least to the writings of Chrysippus (c.282–206 BC). Chrysippus was the third leader of the Stoic School (after Zeno of Citium and Cleanthes) and by many accounts was the master theoretician of the Greek Stoic tradition. Almost all of Chrysippus' writings have been lost, but in the seventh book of *Lives of Eminent Philosophers*, Diogenes claims to quote directly from his book *On Ends*. I quote the relevant passage in full, following Hicks's translation:

An animal's first impulse, say the Stoics, is to self-preservation, because nature from the outset endears (*oikeiouses*) it to itself, as Chrysippus affirms in the first book of his work, *On Ends*: his words are, 'The dearest thing (*proton oikeion*) to every animal is its own constitution and its consciousness thereof'; for it was not likely that nature should estrange the living thing from itself or that she should leave the creature she has made without either estrangement from (*allotriosai*) or affection for (*oikeiosai*) its own constitution. We are forced then to conclude that nature in constituting the animal made it near and dear to itself (*oikeiothai pro eauto*); for so it comes to repel all that is injurious and give free access to all that is serviceable or akin to it (*ta oikeia*).¹⁵

Both in the direct quotation from Chrysippus and in the attendant commentary from Diogenes we can see the basic outlines of the doctrine of *oikeiosis*. As presented here it combines cosmological and psychological commitments. The cosmological doctrine involves a view of animals as the products of nature (*physis*) which is itself conceived to be a unified and rational creative force. The basic psychological doctrine is a form of what we would now call psychological egoism: animals (including humans) are said to be endowed with an innate impulse or instinct (the Greek is *horme*, root of the modern

¹⁴ Pembroke 1971, 115.

¹⁵ Diogenes Laertius, *Lives* 7.85; tr. R. D. Hicks. Plutarch reports a similar claim as appearing in book 1 of Chrysippus' lost work *On Justice*; see Plutarch, *On Stoic Self-Contradictions*, *Moralia* 1038c.

biological term, 'hormone') toward self-preservation. Hence whenever an infant suckles or a turtle struggles to right itself (two of the stock examples) Stoics see *oikeiosis* at work: a rational natural order is so constituted to ensure that animals are immediately drawn toward what serves and preserves them.

Two features of this early treatment of the doctrine deserve comment. First, although I have been talking so far of this whole complex of claims as the Stoic doctrine of *oikeiosis*, it is worth taking note of the particular ways in which the term (and its grammatical relatives) figure in the articulation of the theory. The concept enters first in stating the relationship that nature establishes between an animal and itself, or more specifically between an animal and its constitution (*sustasis*). In Hicks's translation this is said to be a relationship of endearment (*oikeiotes*), and the animal's constitution is said to be 'the dearest thing' (*proton oikeion*). With an ear for the etymological origins we might think of this as the animal being innately 'at home' in its own body. In keeping with the inferential pattern already noted, this preestablished 'endearment' is here contrasted with a state of alienation or estrangement (*allotriosai*). Notice, however, that the *oikeiosis* said to be manifested in an animal's self-concern immediately reappears in its relations to various objects and states in its environment. Some of these objects are now encountered as *ta oikeia*: things that are 'appropriate' (or 'serviceable and akin', as Hicks has it) to the animal's preservation.¹⁶ These are contrasted to those that are injurious or in some way threaten harm. *Oikeiosis* in this way begins as a self-relation but immediately broadens to inform and structure the animal's experience of the objects around it.

The second point to note here is that Chrysippus' doctrine involves an appeal to some sort of self-consciousness or self-understanding. This thought is already to be found in that portion of the passage that Diogenes claims to derive directly from his source. In Hicks's rendering: what is dear to the animal is both the animal's own constitution and its consciousness thereof.¹⁷ We must take care not to interpret this claim prematurely. The Greek term is *syneidesin*, which we might transliterate as 'occurring with ideas'. Hicks's translation can certainly be justified on etymological grounds; the Latin roots *con-scio* literally mean 'with-knowing', and hence echo the etymology of the Greek. But we must here check our tendency to import anachronistic preconceptions about what this consciousness amounts to. For now we can simply note that already among the earliest statements of the *oikeiosis* doctrine it is associated with some kind of self-awareness or self-understanding. I return in section 4 to consider just what this 'with-knowing' might consist in. It is also worth noting that we here encounter a second Greek term that is absent from the Platonic and Aristotelian corpus. It is found,

¹⁶ Elsewhere in the same passage Hicks renders this important phrase with the rather obscure dietary term *aliments*: that which nourishes or feeds.

¹⁷ Long and Sedley render this sentence as follows: 'The first thing appropriate (*proton oikeion*) to every animal . . . is its own constitution and the consciousness of this.' Long and Sedley 1987, i. 346.

however, in the Greek of the New Testament and the early Christian church, where it is standardly translated as 'conscience'.¹⁸

One final point is worth adding before turning to the Stoic deployment of *oikeiosis* in disputes and proofs. At least since later Greek antiquity, and quite probably earlier as well, the doctrine of *oikeiosis* has been associated with an image: that of a set of concentric circles.¹⁹ In some recent discussions the circles have been associated specifically with what is distinguished as 'social *oikeiosis*' as opposed to 'individual *oikeiosis*',²⁰ but there is reason to believe that these are best understood as two aspects of a single unified view. The central circle is identified with the individual, with progressively broader circles marking the domains of the immediate family, household, city, and so on. The broadest circle is associated with the whole of humanity, or of rational beings in total. The image of the circles requires interpretation, but it suggests that *oikeiosis* is not to be understood simply as a psychological state or disposition but as a process. As a pebble dropped in water creates a spreading set of circles, so in psychological maturation the self-concern at work in *oikeiosis* tends systematically to broaden its scope to encompass not just the individual but a progressively larger domain of those around him. We have just seen one example of this broadening in the transfer of *oikeiosis* from an animal's self-concern to a concern with objects in the immediate environment. As we shall see presently, the Stoics also made a much more contentious claim about the form this broadening takes in specifically human maturation.

2. *Oikeiosis* and Stoic Cosmopolitanism

I turn now to consider the role played by *oikeiosis* in two important Stoic arguments. The first of these concerns the Stoics' distinctive commitment to cosmopolitanism. The term 'cosmopolitan' is said to derive from a saying of the Cynic Diogenes,²¹ but it was the Stoics, among the ancient schools, who did most to promote the cosmopolitan ideal. A passage from Plutarch's book on Alexander the Great exhibits key elements of Stoic position.

The much admired *Republic* of Zeno . . . is aimed at this main point, that our household arrangements should not be based on cities or parishes, each one marked out by its own legal system, but

¹⁸ The term occurs quite regularly in the Pauline epistles. A particularly important example comes in Romans 2: 15: 'They show that what the law requires is written on their hearts, to which their own conscience (*syneideseos*) also bears witness . . .' I am grateful to David McNeill for his assistance in researching this textual history.

¹⁹ On Hierocles' use of the image, see later in this chapter. Cicero uses the image at *De Officiis* 1.54, a text which is known to have exercised considerable influence on Kant. The image of the circles became common in Enlightenment discussions of sympathy.

²⁰ See in particular Julia Annas's edition of Cicero's *De Finibus, On Moral Ends*, tr. Raphael Woolf (Cambridge: Cambridge University Press, 2001), 69 n. 7. The distinction between individual and social *oikeiosis* is emphasized in the surviving portions of Hierocles' *Elements of Ethics*, of which more later.

²¹ See *Lives* 6.63. When asked where he was from Diogenes reportedly would answer: 'I am a citizen of the world (*kosmopolites*).'

we should regard all men as our fellow citizens and local residents, and there should be one way of life and order, like that of a herd grazing together and nurtured by a common law. Zeno wrote this, picturing as it were, a dream or image of a philosopher's well regulated society.²²

It is worth distinguishing two dimensions of ethical universalism in the position Plutarch attributes to Zeno. The first is universalism in the scope of moral theory. That is, it is an answer to the question: 'To whom, exactly, do our ethical standards apply?' Here Zeno's answer seems to be: *everyone*. He envisions a single, universal legal code—a 'common law' for all—encompassing not only all 'cities and parishes' but every caste and rank within society as well. The most prominent ancient Stoics famously included an emperor (Marcus Aurelius) and a slave (Epictetus); Stoic moral teachings purport to provide rules of conduct equally suited to both.

This first dimension of universalism in Stoic moral theory was to exercise considerable influence on later ethical traditions (starting already with early Christianity), and it certainly marked a break from the ethical systems that prescribed rules of conduct only for those of a particular faith or community. But among the ancient philosophical schools this alone cannot be said to mark a fundamental break, even if it was given new emphasis and prominence by the Stoics. For in retrospect we can recognize elements of this sort of ethical universalism in the positions of Plato and Aristotle, particularly insofar as their ethical theories were rooted firmly in accounts of human (as opposed to narrowly Greek or Athenian) psychology. Nonetheless, there is no denying that there were crucial limitations of scope in pre-Stoic Greek ethics (Aristotle's account of slavery being the most notorious example), so the Stoics could rightly claim to have advanced a universalism that was at most incipient in earlier moral theory.

But Stoic cosmopolitanism also made a more radical break—and occasioned much more controversy—with a second claim to universality. In Plutarch's report we see this second form of universalism at work in the claim that 'we should regard all men as our fellow citizens and local residents'. Here the question is not only one about the scope of our ethical *theory*; it is an issue about the scope of our ethical *commitments*. In contrast to virtually all earlier moral traditions, which assumed that the sphere of justice was delimited (whether to one's family, one's tribe, one's co-religionists, one's city, or to those with whom one had some direct contact), the Stoics held that the sphere of moral concern must in the limiting case extend to all rational beings. As Plutarch was writing, this 'philosopher's dream' was already implicated in the emergence of ancient imperialism, first in its Greek and then much more systematically in its Roman manifestations. But at its core lay a novel and controversial ethical ideal.

Stoic cosmopolitanism has received considerable attention in recent years, in no small part because cosmopolitanism itself has been an intensely disputed ideal in recent times.²³ But what matters for my purposes here is the role played in motivating it

²² Plutarch, *On the Fortune of Alexander* 329A–B.

²³ See inter alia, Nussbaum 1997, 2000; Pangle 1998; Padgen 2000; Hill 2000; Berges 2005.

by the doctrine of *oikeiosis*. An excerpt from Hierocles in Stobaeus' anthology will help bring out the connection. Hierocles begins with an invocation of the image of concentric circles, here elaborated in considerable detail:

Each one of us is as it were entirely encompassed by many circles, some smaller, others larger, the latter enclosing the former on the basis of their different and unequal dispositions relative to each other. The first and closest circle is the one which a person has drawn as though around a centre, his own mind. This circle encloses the body and anything taken for the sake of the body. For it is virtually the smallest circle, and almost touches the centre itself. Next, the second one further removed from the centre but enclosing the first circle; this contains parents, siblings, wife, and children. The third one has in it uncles and aunts, grandparents, nephews, nieces, and cousins. The next circle includes the other relatives, and this is followed by the circle of local residents, then the circle of fellow-tribesmen, next that of fellow citizens, and then in the same way the circle of people from neighboring towns, and the circle of fellow-countrymen. The outermost and largest circle, which encompasses all the rest, is that of the whole human race.²⁴

The very structure of Hierocles' image might already be taken to suggest a step toward cosmopolitanism, insofar as it involves situating the local community in relationship to the whole of humanity in the same sort of relationship that holds e.g. between a family and the city of which it forms a part. This implicit cosmopolitan orientation soon becomes explicit in the form of an ethical prescription:

Once all these [circles] have been surveyed, it is the task of a well tempered man, in his proper treatment of each group, to draw the circles together somehow toward the center, and to keep zealously transferring those from the enclosing circles into the enclosed ones.

Among other things, Hierocles' fragment exhibits some of the distinctive texture of Stoic cosmopolitanism. Hierocles, at least, does not seem to envision a kind of flat moral universalism, in which my obligations to those distant in time and space somehow equals (or ultimately overshadows) the special commitments I have to those in my family, or my village, or my academic community. Nor does Hierocles rule out the possibility that Plutarch's Zeno seems to exclude, namely that my deliberations might in some important sense be 'based at home'. The picture Hierocles offers is one where the distinction between near and far is still relevant to my deliberations; his point is that the scope of what is ethically relevant is universal, and that I have an obligation to 'draw nearer' those who start out far away. Hierocles even specifies just how much closer the distant should be brought.

It is incumbent upon us to respect people from the third circle as if they were those from the second, and again to respect our other relatives as if they were those from the third circle . . . The right point will be reached if, through our own initiative, we reduce the distance of the relationship with each person.

²⁴ Hierocles, fragment excerpted in Stobaeus' *Eclogae* 4. 671ff.; Long and Sedley 1987, i. 349.

This feature of Hierocles' cosmopolitanism has been overlooked by some influential recent commentators, who have tended to treat all Stoic cosmopolitanism on the model of the position attributed to Zeno.²⁵

What does all this have to do with *oikeiosis*? The answer should initially surprise us. Up to this point we have located the core of the *oikeiosis* doctrine in a form of psychological egoism, but in the debate over cosmopolitanism the concept is used to justify a very robust account of the demands of justice. How could a form of self-interest or self-concern serve as the basis for such an expansive and unprecedented ethical demand? It is crucial to see that the Stoics' answer is *not* a precursor of the standard modern social contract approach, imagining a group of self-interested individuals bargaining over the rules for a just society. Rather, their position seems to be that cosmopolitan concern is the final stage in the process of *oikeiosis*, as the self-concern already at work in the infant systematically expands as part of the natural process of maturation. This would seem to be a difficult claim to sustain, given the prevalence of xenophobia and other forms of parochialism in human existence as we actually observe it. But the Stoics held that such limitation of ethical perspective is properly understood as resulting from a disruption or corruption of the natural process whereby one's sphere of concern grows progressively wider. Only in the Sage, perhaps, does it reach its widest and most fully developed extent, but that breadth of moral concern is the natural outcome of a natural process; it should therefore orient us in the cultivation of our moral instincts and capacities for deliberation, and even (if one happens to be a Stoic emperor) in one's political and military endeavors.

Like almost every Stoic teaching, this claim about the foundations of justice has been subject to severe criticism, both by ancient and modern critics, and it is easy to see that it is beset by a number of serious difficulties. In the ancient world this was thought to be another of those repugnant Stoic paradoxes, and was widely held to be just as preposterous as the claim that e.g. health does not contribute to happiness. The idea that one might care equally for a Greek and 'the most distant Mysian' struck many of the critics of Stoicism as not only straightforwardly false but also pernicious. After all, to abolish the idea that members of my family (or my fellow citizens or my academic colleagues . . .) have special claims upon my action would effectively be to abolish the family (and the state and academies . . .) as morally significant institutions. And this was rightly seen as quite antithetical to justice. It is in one such refutation—this from a late skeptical textbook only recently recovered—that we find one of the clearest statements of the inferential role of *oikeiosis* in the Stoic cosmopolitan argument.

²⁵ See in particular Nussbaum 1997, 7: 'This being so, Stoic cosmopolitans hold, we should regard our deliberations as, first and foremost, deliberations about human problems of people in particular concrete situations, not problems growing out of a local or national identity that confines and limits our moral aspirations.' This may be true of the position reported by Plutarch, but Hierocles seems explicitly to allow for a form of cosmopolitan deliberation that 'grows out of a local . . . identity', and even, by the specified degree, 'limits our moral aspirations'.

We have an appropriate relationship to members of the same species. But a man's relationship to his own citizens is more appropriate. For appropriation varies in its intensification. So [as regards] those people [the Stoics] who derive justice from appropriation: if on the one hand they are saying that a man's appropriation to himself is equal to his appropriation in relation to the most distant Mysian, their assumption preserves justice; on the other hand, no one agrees with them that the appropriation is equal. That is contrary to plain fact and one's self-awareness . . . If on the other hand they themselves should say that appropriation can be intensified, we may grant the existence of philanthropy, but the situations of two shipwrecked sailors will refute them.²⁶

Notice that variants of the term *oikeiosis* (appropriation) occur eight times in this densely argumentative fragment. The unknown skeptical author explicitly names the inferential pattern we have been discussing: the Stoics claim to 'derive justice from *oikeiosis*'. But this derivation falters, according to the critic. The Stoics claim that *oikeiosis* generates a concern that extends to the whole of humanity, but such a kinship is said to be insufficient to 'preserve justice'. The refutation turns on a dilemma concerning variability in *oikeiosis*. If the strength of species-wide *oikeiosis* is held to be strictly equivalent to that of narrowly self-directed *oikeiosis*, then the Stoic psychological thesis is 'contrary to plain fact'. But if even the slightest degree of variation is admitted then a conception of justice derived from *oikeiosis* has morally objectionable consequences and fails in its claim to warrant a cosmopolitan ethic. The explication of the argument is not fully spelled out in the portion of the text that has survived, but it is not hard to extrapolate the 'case of the shipwrecked sailors' which is meant to press it. Philanthropy may indeed be natural, but the circumstances of justice make themselves felt precisely in those situations where the self-interest of one agent comes into conflict with the interests of others. If two sailors are clinging to a timber that suffices only for keeping one afloat, then one or the other will have to drown. But if one's natural concern for one's own preservation is greater by even the smallest fraction than one's concern for the wellbeing of others then Stoic justice would seem to require that one always chooses self-preservation at the expense of others in such circumstances.²⁷

It is certainly tempting to propose a Stoic reply to this criticism, but at this point I forego further descent into the dialectic. For we have gone far enough to extract what we need concerning the inferential role of the concept at work here. Summing up what we have learned so far we can conclude at least this much: *oikeiosis* is a foundational concept in Stoic psychology and ethics.²⁸ It is held to be a psychological fact of direct

²⁶ *Commentary on Plato's Theaetetus*, 5.18–6.31; Long and Sedley 1987, i. 350.

²⁷ This line of criticism seems to have been pioneered by Carneades. For an analysis see Striker 1991, 50–61.

²⁸ This claim has been challenged in Striker 1983, but her thesis is carefully qualified. Striker does not dispute the attribution of this cosmopolitan argument to the Stoics, with the appeal to *oikeiosis* serving as a premise. Her contention is that the appeal to *oikeiosis* does not serve as the basis for the more fundamental ethical claims that the life lived in accordance with nature is the virtuous life, and that virtue is sufficient to happiness.

ethical significance. *Oikeiosis* underlies and explains an organism's innate concern for its own preservation, but it also expands to incorporate more and more within its appropriated domain. In this sense it involves both a relation (to oneself and to things in one's environment) and a process (whereby the sphere of concern expands). Properly cultivated, *oikeiosis* can progressively expand to incorporate the whole human domain. From the outset it provides normative guidance in action, informing an appreciation for the difference between harmful and beneficial endeavors. It involves a distinctive form of 'being-at-home' or familiarity (as contrasted to a state of alienation or estrangement) with oneself and with one's environment. And it is used to motivate a form of cosmopolitanism in which one is held to be properly 'at home' in the human sphere as a whole.

3. *Oikeiosis*, Pleasure, and Desire

I turn now to a second Stoic argument, and at the same time from Greek to Roman Stoic sources. Our text in this instance is Cicero's *De Finibus*; the argument emerges in the context of Stoic attempts to refute Epicurean moral psychology. This dispute is of considerable interest in its own right, and will help to fill out our understanding of the texture of the Stoic position. But it is also particularly important for my purposes, since we here begin to see traces of the transcendental dimension of Stoic argumentative strategy.

Cicero himself was not a Stoic; indeed he was among the fiercest critics of Stoic teachings. Nonetheless his account of Stoic doctrine in *De Finibus* remains one of the most complete extant statements of Stoic moral philosophy, and it seems to have been composed with the benefit of direct knowledge of a number of Greek Stoic sources that have now been lost. Its importance also derives from the fact that in it Cicero systematically and quite deliberately undertakes the task of rendering Greek philosophical terminology in Latin. *De Finibus* is constructed as a series of dialogues in which the teachings of each of the predominant Hellenistic schools are systematically expounded and assessed. Books 3 and 4 are devoted to Stoic teachings, with Cato in the role of enthusiastic expounder of Stoicism while Cicero himself acts first as cooperative audience and then as vehement critic. The third book contains Cato's exposition of Stoicism; book 4 develops Cicero's refutation.

The context in which the doctrine of *oikeiosis* first appears in *De Finibus* is significant. Book 3 opens with some stage-setting and unsystematic sparring about Stoic doctrines, particularly as regards Stoic claims to originality and regarding the core Stoic claim that virtue is sufficient to happiness. However, Cato soon proposes that a more systematic exposition is required—a proposal to which his interlocutor readily agrees. The narrative voice says simply 'He began,' and there follows Cato's statement, now in Ciceronian Latin, of Stoic theory. I recount the dramatic set-up because it bears directly upon our topic; for the very first concept Cato introduces is the notion of *oikeiosis*:

[Cato:] It is the view of those whose system I adopt that immediately upon birth (for that is the proper point to start from) a living creature feels an attachment to itself, and an impulse to preserve itself and to feel affection for its own constitution and for those things which tend to preserve that constitution; while on the other hand it conceives an antipathy to destruction and to those things which appear to threaten destruction.²⁹

So far this is familiar territory; indeed it is effectively a reiteration of the doctrine Diogenes had reported from Chrysippus, albeit now translated into a new Latin vocabulary. *Oikeiosis* is here rendered with an elaborate phrase: the organism is ‘accordant with itself’ [*sibi conciliari*] and ‘commended to its own preservation’ [*commendari ad se conservandum*]. And what in the Greek had been simply *ta oikeia* is here ‘things which tend to preserve one’s constitution’ (*eaque quae conservantia sunt eius status*). There may seem to be a difference in the scope of the thesis: Chrysippus had applied it to animals but the text here refers to ‘all living creatures’. But this is an artifact of Rackham’s translation; the Latin is simply ‘*animal*’. Certainly the core claim as to the innateness of the principle remains the central emphasis.

But as Cato’s exposition proceeds we encounter a deployment of *oikeiosis* in an inferential context that we have not yet considered. Having stated the Stoic doctrine, Cato immediately sets out to justify it.

In proof of this opinion they urge that infants desire things conducive to their health and reject things that are the opposite before they have ever felt pleasure or pain; this would not be the case, unless they felt an affection for their own constitution and were afraid of destruction. But it would be impossible that they should feel desire at all unless they possessed self-consciousness, and consequently felt affection for themselves. This leads to the conclusion that it is love of self which supplies the primary impulse to action.

Cato’s proposed proof is tantalizing but also frustratingly underdeveloped; Long has called it ‘an argument of lightening brevity’.³⁰ The first thing to note is its dialectical context. Cato sets out to prove his thesis not from first principles but in a kind of determinate negation of its main contemporary rival. That rival, of course, is the core thesis of Epicurean psychology: the claim that the overarching motive in human endeavor is the desire for pleasure. Cato tackles this claim in the context of what Brunschwig has called ‘cradle arguments’—the disputes between the two schools as to the psychological traits of newborns, whom Epicurus himself had famously described as ‘mirrors of nature’.³¹

How is Cato’s proof to be reconstructed? The answer is far from clear. At least part of the argument seems to turn on what we would now call an empirical claim. In his attempt to forestall the Epicurean claim that we are all born pleasure-seekers, Cato seems to deny that newborn infants experience pleasure or pain. This is, to say the least, a surprising claim by modern lights, and it is tempting to dismiss it as a historically revealing but philosophically uninteresting artifact of ancient attitudes about

²⁹ Cicero, *De Finibus* 3.16.

³⁰ Long 1993, 254.

³¹ Brunschwig 1986.

babies. But this would be to miss something important. Notice first that Cato's denial of newborn pain and pleasure is framed in a comparison: prior to the experience of pleasure or pain comes 'the desire for things conducive to their health'. This may not be enough to win us over to Cato's neonatological thesis, but we should not mistake it for an image of newborns as somehow less-than-fully alive. On the contrary, newborns in Cato's account are already quite discriminating agents.

Elements of Cato's reasoning come into focus when read in light of an image found in the Greek Stoic sources: the image of pleasure as a flower or bloom which makes its appearance only after an organism has satisfied its needs. The infant, to take up the lead from this metaphor, would never have the good fortune to experience pleasure unless it was endowed with a disposition which leads it to fulfill the vital needs from whose satisfaction pleasure 'blossoms'. The contention against Epicureans is that a nested set of psychological conditions must already be in place before pleasure can be encountered and found desirable. The basic condition, already immediately at work in the newborn, is the feeling of affection for and implicit understanding of the organism's own constitution and what preserves it. This is the core state of *oikeiosis*. This state manifests itself in the infant's disposition to seek things conducive to its health and to eschew the opposite. It is only as an effect of all this that the infant comes to feel pleasure—as an outcome of a causal sequence stemming from this underlying state. Hence the surprising thesis: infants do not experience pleasure *ab initio*, but encounter it only as a pleasurable downstream effect of a prior motivational condition. In the case of infants this whole process may unfold very quickly, as the first instincts very quickly produce the first pleasures. But the key point for Cato is that the instinct must come first, with the pleasure to follow only as its dependent effect. A version of this line of argument can be found in Diogenes, and it is in close keeping with Cicero's own attack on Epicurean psychology in *De Finibus* 2.³²

So interpreted, Cato's argument against the Epicureans is certainly not without interest, but it is not without problems either. It is natural to wonder, for instance, whether an infant might encounter its first pleasures as a matter of good fortune, or under the guidance of some beneficent carer. In that case the child might experience its first pleasures as a kind of psychic onslaught, rather than as the 'fruit' or 'bloom' of prior motivational states, and hence without the sort of motivational

³² For a Greek statement of this argument see Diogenes, *Lives* 7.86: 'As for the assertion made by some people that pleasure is the object to which the first impulse of animals is directed, it is shown by the Stoics to be false. For pleasure, if it is really felt, they declare to be a by-product, which never comes until nature by itself has sought and found the means suitable to the animal's existence or constitution; it is an aftermath comparable to the condition of animals thriving and plants in full bloom.' Cicero's rejection of Epicurean hedonism about infants is stated as follows: 'Again in the infant the natural instinct is not to seek pleasure; its instinct is merely towards self-regard, self-preservation and protection from injury. Every living creature, from the moment of birth, loves itself and all its members; primarily this self-regard embraces the two main divisions of mind and body, and subsequently the parts of each of these. Both mind and body have certain excellences; of these the young animal grows vaguely conscious, and later begins to discriminate, and to seek for the primary endowments of Nature and shun their opposites' (Cicero, *De Finibus* 2.33).

prerequisites Cato asserts. This line of objection becomes much more pressing when we remember that Cato's argument was meant to apply not just to pleasure but to pain. Does the infant really need to be 'afraid of destruction' in order to experience pain? What about the pain the infant encounters in the process of childbirth itself? Is there any reason to think that this presupposes a prior motivational structure in the not-yet-newborn? It seems entirely open to the Epicurean to claim that an infant's very first experience is an experience of pain, and its first desire is the desire to be rid of it!

Rather than considering Stoic replies to these objections, I want here to take note of a formal feature of Cato's argument. Notice in particular the way in which his proof moves from an undeniable or universally accepted psychological premise (the claim that we feel pleasure) to a contentious and substantive conclusion (the doctrine of *oikeiosis*). The route from premise to conclusion is an argument from the accepted fact to its underlying condition. In this case the trajectory of argument is temporal and causal. That is, Cato argues from universally acknowledged psychological effect to the prior causal conditions that bring it about. But note also that we have not yet taken the full measure of Cato's proof. To this point we have focused on his first claim, namely, that infants experience pleasure and pain only after desiring things conducive to their health. Recall that Cato goes on to advance a second claim. He argues that desiring itself would be impossible without self-consciousness and self-affection: '*it would be impossible that they should feel desire at all unless they possessed self-consciousness (sensus sui), and consequently felt affection for themselves*'. In the context of his overall argument, this seems to be a case of drilling down from condition to condition. The first stage of the proof argued that the experience of pleasure presupposes desiring things conducive to health; the second stage now argues that desiring in turn presupposes a form of self-consciousness. Cato deploys the two claims in tandem, but in principle the second claim could stand alone as an argument about the conditions on the possibility of an undeniable or universally accepted premise: that we feel desire. Before we can assess this second claim we must squarely face a question that we have thus far been carefully deferring: *what exactly did the Stoics mean by self-consciousness?*

4. Seneca on Animal Self-Consciousness

In our encounters with the Stoic notion of *oikeiosis* we have found it to be developed in close connection with claims about self-consciousness or self-awareness—*syneidesis* in Chrysippus, *sensus sui* in Cicero. Having developed a firmer grip on the concept of *oikeiosis* itself, we must now tackle this issue directly. What kind of self-consciousness did the Stoics hold to be at work in *oikeiosis*? In modern times, our thinking about self-consciousness has been decisively shaped by the Cartesian/Lockean tradition which conceives of self-consciousness as a kind of 'internal' psychic self-presence, a

distinctive awareness we have of our own occurrent psychological states.³³ Over the last century or so, the traditional authority of self-consciousness has been challenged from virtually every quarter.³⁴ But even where self-consciousness has been attacked, it is typically a version of Cartesian self-consciousness that is presupposed among the critics. In tackling the Stoic claims about self-consciousness we must take care not to import our characteristically modern preconceptions about what is under discussion.

In a recent discussion of these matters, Long has argued that Stoic self-consciousness is aptly understood in terms of the modern notion of proprioception, the quasi-perceptual awareness an organism has of the bearing of its own body.³⁵ There are some obvious advantages to this line of interpretation: it renders somewhat more palatable the otherwise surprising claim that self-consciousness is present in all animals; and it provides the distinctive form of legitimacy that comes of finding a respectable modern equivalent for an ancient doctrine. But while Long's thesis is not all wrong, it is not all right either. Or so I shall argue here. To make out this case I turn to a relatively late Stoic source: the letters of Seneca. Seneca's Stoicism is far less systematic than that we find articulated in Cicero or even in the fragments from Stobaeus. (The dust-jackets of the Loeb edition describe his writings as 'more clever than profound'.) But for our purposes he is a crucial figure, because he provides the most direct and extensive consideration of the Stoic thesis that *oikeiosis* is or involves self-consciousness.

By tradition, Seneca's 121st letter to Lucilius carries the title, 'On Instinct in Animals', although in Seneca's original it bears only the heading: '*Seneca Lucilio suo salutem*'. Seneca's letters take many forms: some are meditations, some exhortations to Lucilius on some moral matter or another, a few take the form of abstract philosophical expositions. But the letter on instinct recounts and revisits a dispute. After two paragraphs of preliminaries, the disputed thesis comes out into the open: 'We were once debating whether all animals had any feelings about their constitution.'³⁶ Seneca himself describes this as a 'little question' (*quaestiunculam*), in contrast to the correspondingly large ethical questions over which Lucilius has reportedly been pressing him ('Prove to me that felicity is fickle and empty'; 'How

³³ I don't mean here to take a stand on the proper interpretation of the texts of Descartes or Locke. In referring to the 'Cartesian/Lockean' conception of self-consciousness I have in mind the conception that has been commonly attributed to them by commentators and critics alike. For an argument that Descartes did not himself have a 'Cartesian' conception of self-consciousness, see Balibar 1998.

³⁴ A few examples that could easily be multiplied: Ryle and Heidegger rejected, from quite different positions but for similar reasons, the idea that self-conscious subjectivity is the hallmark of human existence. Semantic externalism challenged the assumption that individuals have a privileged knowledge of the content of their own thoughts or the meaning of their utterances. Behaviourism, functional state identity theory, and psychoanalytic theory each in their own way challenged the traditional assumption that to be in a state of mind is *ipso facto* to know that one is in that state. And countless empirical studies seemed to tell us that we are often quite ignorant as to our own psychological states.

³⁵ Long 1993.

³⁶ Seneca, *Epistles* 121.5; emphasis added. I'm not entirely happy with Gummere's translation. Here is the Latin: *Quaerebamus, an esset omnibus animalibus constitutionis suae sensus?* More literally: 'We were asking whether every animal has a sense for its constitution.'

can I crave less and fear less?'; . . .). But Seneca also hints that this little question may be of considerable significance, pertaining to 'the nature and origin of character'.

Seneca's own answer to the disputed 'little question' is clearly affirmative and is stated explicitly at the outset of the fourth paragraph: 'So all these animals have a consciousness of their physical constitution . . .'³⁷ The precise scope of this claim is not entirely clear, but over the course of the letter Seneca discusses cats, hawks, chickens, peacocks, turtles, bees and spiders, including thereby not only mammals, birds, reptiles, and insects, but both some animals long celebrated for their intelligence and others notorious for their stupidity. In the main body of the letter Seneca defends his thesis, first by proposing a positive proof, then with replies to a series of objections. The positive proof is grounded in observations about the behavior of animals. The three replies in turn dispatch an Epicurean explanation of the same phenomenon, reply to the charge that Stoics have overintellectualized animal and child behavior, and deal with a challenge pertaining to personal identity over time.

To modern ears, Seneca's arguments curiously combine prescient scientific hypothesis and striking non sequitur. In his reply to the Epicureans, Seneca argues forcefully and resourcefully against attempts to explain animal behavior as entirely learned from experience. On the contrary, Seneca argues, core elements of that behavior must be acknowledged as both innate and species-specific. The affinity of Seneca's position on this point with modern biological accounts has not escaped attention. Gummere calls Seneca's nativist thesis 'sound and modern'; Long and Sedley treat it as 'an attempt, and an interesting one, to do justice to data which would now be explained by reference to natural selection and genetic coding.'³⁸

But Seneca's positive argument will seem much less compelling to a modern audience. His main positive evidence for animal self-consciousness is the 'apt and expedient' bodily skills animals exhibit, an agility he compares to some stock examples of skillful behavior in the crafts:

That this is the case is proved particularly by their making motions of such fitness and nimbleness (*apte et expedite movent*) that they seem to be trained (*erudite*) for the purpose. Every being is clever (*agilitas est*) in its own line. The skilled workman handles his tools with an ease born of experience; the pilot knows how to steer his ship skillfully; the artist can quickly lay on the colors which he has prepared in great variety for the purpose of rendering the likeness, and passes with

³⁷ *Ergo omnibus constitutionis suae sensus est . . .* Seneca, *Epistles* 121.9; emphasis added. It is worth registering an observation about the syntax of Seneca's *sensus*, which I find myself hard-pressed to explain. Gummere's translation has Seneca asking whether animals 'have a feeling' and answering that they 'have a consciousness'. The noun is the same in both cases: *sensus*. But Seneca's verb in each case is a form of *esse*, to be. So it would perhaps be a bit more literal to cast his thesis thus: 'Everything is sensible of its own constitution.' But *sensus* is a noun. I can find no easy English equivalent for this syntactical form.

³⁸ Gummere (editor and translator), *Seneca's Epistles 93–124* (Loeb Classical Library, 1925, 2000), 406n.; Long and Sedley 1987, i. 351.

ready eye and hand from palette to canvas. In the same way an animal is agile in all that pertains to the use of its body.³⁹

Seneca's claim seems to be that the similar phenomena must have a similar cause. Since artisan agility is the product of intimate knowledge of the tools and materials that are so aptly handled, Seneca concludes that animal agility must be an expression of an analogous familiarity with that which animals handle so adeptly—that is, their own bodies. The difference is that, while an artisan's knowledge of his tools is acquired through training and experience, an animal's bodily self-knowledge is part of its innate endowment:

But that which art gives to the craftsman, is given to the animal by nature. No animal handles its limbs with difficulty, no animal is at a loss how to use its body. This function they exercise immediately at birth. They come into the world with this knowledge (*scientia*); they are born full-trained.⁴⁰

To rely on this line of argument is to find oneself squeezed from both sides of modern debates about self-consciousness. Cartesians will certainly deny that any behavioral criterion could suffice to prove self-consciousness, which is something that can be established only by and for the self-conscious being itself. No amount of animal agility could suffice to demonstrate that animals have an inner conscious life. But Descartes's modern critics will find the Stoic argument no more palatable, since they will simply deny that self-consciousness or self-awareness is needed to explain 'apt and expedient' animal behavior. Such behavior can be assumed to be hardwired for survival, without requiring any kind of psychic self-presence. So from both sides of this familiar modern divide, Seneca's argument looks to be a plain non sequitur. Before putting this down to Seneca's failings as a philosopher, however, we should once again insure against anachronism by deploying our inferentialist approach. If Seneca confidently relies on an argument that is plainly fallacious when interpreted in modern terms, then perhaps this indicates that he means something quite different by 'self-consciousness' than we moderns do.

So what, exactly, does Seneca mean to attribute to animals? The closest Seneca comes to addressing this issue directly comes in his reply to the second objection, which charges that the Stoics have overintellectualized animal psychology. Interestingly, the turf on which the dispute unfolds shifts at this point from animal to child psychology.

But some object as follows: 'According to your account, one's constitution consists of a ruling power in the soul which has a certain relation towards the body. But how can a child comprehend this intricate and subtle principle, which I can scarcely explain even to you? All living creatures should be born logicians, so as to understand a definition which is obscure to the majority of Roman citizens!'⁴¹

In defending the Stoic thesis from this objection, Seneca sets out to clarify the self-conscious self-knowledge that is at issue. The first crucial point to notice is the

³⁹ Seneca, *Epistles* 121.5–6.

⁴⁰ Seneca, *Epistles* 121.6.

⁴¹ Seneca, *Epistles* 121.10.

specification of the *object* of self-consciousness. What is known in self-consciousness is said to be ‘the constitution’ of the organism. This is a feature of the Stoic position that we can see at work throughout the tradition. In the position attributed to Chrysippus, animal self-consciousness is said to be of ‘of its own constitution’ (*hauton sustasis*); Cicero renders this term both as *status* and as *constitutio*, which is the term found in Seneca. Seneca’s text explains what is meant by this. An animal’s constitution is said to be its ruling power or principle. It is that part of the animal’s nature that determines how it is to behave, how its parts interact properly, and ultimately how it is to preserve itself and thrive. Hence in attributing innate self-consciousness to all animals, the Stoics are attributing a form of *practical self-comprehension*. Already with this first point we can mark the first fundamental difference between Stoic and Cartesian self-consciousness. Whereas Cartesian self-consciousness first and foremost yields a knowledge of my existence, and specifically of my psychological existence, Stoic self-consciousness is a form of *bodily comprehension*—not a knowledge *that I exist* but an understanding of *what kind of being I am*, and specifically of the ways in which the parts of my body operate together so as to befit the survival (and reproduction) of the whole. It is in this sense that I shall refer to Stoic self-consciousness as both *normative* and *ontological*. It is normative insofar as it provides me with an orientation in the sphere of my practical endeavors; it is ontological insofar as that orientation is rooted in an understanding of my mode of existence as the kind of being that I am.⁴²

But what kind of understanding is this, exactly, such that it might plausibly be attributed to newborns, children, and animals? In replying to his critic, Seneca is at pains to distinguish sharply between this self-conscious self-understanding and the kind of understanding that might be possessed by a zoologist or a psychologist. By comparison to an explicit scientific specification of an organism’s constitution, our innate self-understanding is described as ‘confused, cursory and dark’ (*crasse, summatim et obscure*); it is ‘not very clearly outlined or portrayed’ (*non satis dilucidus nec expressus*). Moreover, it doesn’t provide the animal or child with ‘a definition of his constitution’ (*constitutionis finitionem*), either in the sense of providing explicit definitions of the sort that one would expect in a zoological theory or in the sense of establishing the precise limits of its kind. In short, it is an *implicit and inarticulate self-understanding*, in contrast to the explicit account one might aim for in science.

But there is a further mark of this distinctive self-comprehension that will be particularly important to us in what follows, and this pertains to the term both Cicero and Seneca use in describing self-consciousness. It is a *sensus*. As Seneca writes here: ‘he [the animal] does not know what “a living organism” is, but he *feels* (*sentit*) that he is an animal’. The self-knowledge provided in self-consciousness is

⁴² For a discussion of ontological self-consciousness in a rather different context, see Martin 2010.

not what a later tradition would call a discursive representation; its medium is not conceptual. Indeed Seneca insists that neither the animal nor the child understand what a constitution is, though they do understand their own constitutions. In other words, animals and children are said to lack the concept of that which they comprehend. It is in their attempts to name this distinctive form of *non-conceptual knowledge* that the Roman Stoics use the term *sensus*. We might get at this by saying that the medium of self-consciousness is not ideas but sensations or feelings, but here again we must tread carefully. For Stoic *sensus* must not be confused with the sensations and impressions of the later empiricist tradition. It should be clear that it is not a bare conscious content—what is nowadays called a *qualia*. For it has both a cognitive and a practical dimension that mere sensations lack. Neither is it a narrowly psychological state, if this is meant to be contrasted to a state of one's body. Stoic self-sense is a discriminating bodily self-comprehension that immediately guides and controls one's actions and provides orientation and agility in one's endeavors.

In sum, we can say that the self-consciousness Seneca attributes to animals and young children is an implicit, non-discursive understanding of one's own body and of that which preserves it, informing an ability to find normatively significant differences among the various available courses of endeavor. How does this conception of self-consciousness compare to later accounts? It should be clear from this that the self-consciousness that figures in Seneca's Stoicism differs down the line from that which concerns Descartes. It is not explicit or discursive self-representation but a comprehending sentiment of oneself; it is not knowledge of one's own psychological state but of one's bodily constitution; it is not narrowly factual but immediately normative. The claim is not that animals enjoy a conscious self-presence or an inner life, but that they have an innate self-comprehension that provides them with an implicit answer to the question, 'what kind of being am I?' And its main role is not to underwrite the knowledge that I exist but rather the understanding of what it is appropriate (or inappropriate) for me to do, given the kind of thing that I am. It is at once a form of *consciousness* and a form of *conscience*.

But it should also now be clear that Stoic self-consciousness is not to be equated with proprioception either, although proprioception might indeed be said to be an essential aspect of this form of self-comprehension, and it does contribute to the characteristic agility and even grace of animal movements, which is just what Seneca set out to explain. But Long is wrong in equating the two. Long defines proprioception as a form of perception, taking his lead from Sherrington's canonical description of 'muscular sensations which contribute to perceptions of the relative flexions or extensions of our limbs.'⁴³ The key point is that proprioception, so defined, involves neither the normative nor the ontological dimensions that we have found to be central to Stoic self-consciousness. In short, proprioception

⁴³ Long 1993, 258, quoting Sherrington 1906, 343.

informs me as to the position of my body; it does not tell me what I ought to do with it. Certainly proprioception does not play the role that Cato had claimed for self-consciousness as a condition on the possibility of desire. For it is plain that trauma victims who have lost proprioceptive capacities due to paralysis are still quite able to experience desire.⁴⁴

5. A Transcendental Role?

It is clear from the extant sources that the Stoics took this distinctive form of self-consciousness to be fundamental in at least two respects. In psychology, as we have seen, bodily self-comprehension was held to structure the first motives of an organism, and to provide thereby the basis for infantile agency. In ethics, the doctrine of self-conscious *oikeiosis* was held to provide the proper starting point in the exposition of Stoic moral teachings. But we have also found traces of yet a third respect in which the Stoics held self-consciousness to be fundamental: as a condition underwriting the very possibility of further psychological states. Having arrived at an interpretation of self-consciousness as the Stoics understood it, are we now in a position to reconstruct this line of Stoic thought?

In tackling this question, we can take our bearing from Cato's argument in *De Finibus* where we first encountered this claim. Recall in particular that in the argument against Epicurean psychology, Cicero has his Stoic spokesman claim that 'it would be impossible that they should feel desire at all unless they possessed self-consciousness, and consequently felt affection for themselves' (Cicero, *De Finibus* 3.16). It is striking the extent to which this formula anticipates the distinctive cocktail of themes associated with the later (Kantian) transcendental tradition. As with Kant, we start here with a generally acknowledged fact of experience—in this case the fact that we feel or experience desire—and proceed to an argument that purports to establish an underlying self-consciousness. The path from premise to conclusion runs from conditioned to condition. In the first part of Cato's argument, the conditions at issue were temporal and causal, but here the argument follows along a modal trajectory: from an actual fact to a condition on its possibility ('It would be *impossible* that . . .'). For our purposes this is the most puzzling feature of the argument. Why should self-consciousness serve as a condition on the possibility of desire?

We should take note of the considerable open water that separates Cato's claim from later assumptions about self-consciousness and desire. We tend to think of desire as a relatively primitive psychological condition, widely shared in the animal world. Self-consciousness, by contrast, we tend to think of as a much higher-order cognitive achievement, perhaps as the exclusive privilege (or curse?) of human nature, or at most shared only with a few other intelligent mammals. From this modern perspective,

⁴⁴ For a gripping account of the phenomenology of proprioceptive disruption, see Cole 1995.

Cato's thesis is simply unintelligible. How could the higher order psychological accomplishment serve as the condition on the possibility of the lower order capacity?

One way in which we might reduce the oddity of the claim would, in effect, make Cato's claim trivial. In the form in which we have been considering the argument so far, it is cast as a claim about the conditions on the possibility of *feeling* desire. Desire is a canonical example of a psychological state, so to *feel* desire involves awareness of my psychological state; *felt desire* is thus trivially a form of self-consciousness. But this turns out to be a false lead. 'Feel' is here an artifact of Rackham's translation; the Latin verb is simply a form of *appetere*: to desire, to have an appetite for. Annas, in her translation of *De Finibus*, renders the crucial passage as follows: 'Neither could it happen that they would *seek anything at all* unless they had self-awareness and thereby self-love.'⁴⁵ But again this raises the question: why should seeking something out, simply *having an appetite* for something, presuppose self-consciousness?

In approaching Cato's thesis we will need to draw on other sources, as Cicero's text says nothing to defend it. Once again Seneca's letter on animals proves useful. Although Seneca does not make exactly the same claim as Cicero's Cato, he does seem to rely on the pattern of argument that here concerns us, from acknowledged fact to underlying modal condition.

[A]ll animals possess a consciousness of their own constitutions (*omnia animalia constitutionis suae sensus est*). For they must necessarily feel this (*Necesse est enim id sentiant*), because it is the same agency by which they feel (*sentiant*) other things also . . .⁴⁶

Note in particular the claim to necessity that figures here. Seneca seems to be claiming not only that animals *do* have a kind of self-conscious self-understanding, but that they *must* have it. In Seneca's argument, however, this necessity is keyed not to desire but to perception. Self-awareness is necessary, according to Seneca, because it is requisite for an organism's awareness of things other than its own body. Consciousness of objects, it seems, would be impossible without consciousness of self. Once again, however, we are left to wonder about the basis for this modal claim.

Here an unusual neo-Stoic Greek source provides a crucial lead. In a fragment from Hierocles' *Elements of Ethics*, discovered at Hermopolis early in the twentieth century, we find an elaboration of this line of thought. I quote the relevant excerpts in full, following Long's reconstruction and translation of the fragmentary source:

It seems right to say a few words about sensation (*aistheseos*). For this contributes to knowledge of the first thing which is appropriate (*tou protou oikeiou*), the subject which we said would be the best starting-point for the elements of ethics. We should realize that as soon as an animal is born it perceives itself (*aisthanetai eautou*) . . . The first thing that animals perceive is their own parts . . . both that they have them and for what purpose they have them, and we ourselves perceive our eyes and our ears and the rest. So whenever we want to see something, we strain

⁴⁵ Cicero, *On Moral Ends*, tr. Julia Annas (Cambridge: Cambridge University Press, 2001), 70.

⁴⁶ Seneca, *Epistles* 121.12.

our eyes, but not our ears, toward the visible object. Therefore the first proof of every animal's perceiving itself is its consciousness (*synaisthesis*) of its parts and the functions for which they were given. (Hierocles, *Elements of Ethics*, 1.34–9, 51–7)⁴⁷

Several of Hierocles' claims here should by now be familiar. Ethical theory should begin with an exposition of *oikeiosis*, which is associated with a form of self-consciousness or self-perception—a self-consciousness that is held to be an innate endowment of every animal organism. What is new here is, first of all, a piece of terminology that Hierocles introduces to name this distinctive self-awareness. What Chrysippus had called *syneidesin* and Seneca named as a *sensus sui*, Hierocles here dubs *synaisthesis*—etymologically: 'with-perception' or 'co-perception'. *Synaisthesis* is of course the etymological root for the modern psychological term 'synaesthesia', the perceptual anomaly whereby stimulus for one sensory modality (e.g. the sight of the setting sun) is experienced in a different sensory mode (e.g. as a taste). But the synaesthesia with which Hierocles is concerned is something quite different—not perception across sensory modalities but rather a perception of self that accompanies every perception of an object, an apperception. Like Seneca, then, Hierocles holds that object-consciousness essentially involves self-consciousness.

But Hierocles' fragment goes further than Seneca's letter in explaining why this should be the case. In the perception of an object, he maintains, the perceiving organism exercises its implicit understanding of its own perceptual organs together with an implicit normative grasp of their proper use. In trying to make out a distant object, as Hierocles puts it, 'we strain our eyes and not our ears.' In this respect our perceptual organs are like the artisan's tools; we require a familiarity with their proper use and their proper objects in order to exercise them effectively. Hierocles goes on in the fragment to compare the perceptual organs to animal weaponry: the bull exhibits its implicit understanding of its weapons and their appropriate use when it lowers its head for a charge. Far from being an anomaly, then, Hierocletian synaesthesia is part of the proper functioning of our sensory apparatus. Once again we can think of this self-consciousness as a kind of prior orientation that is brought to the perceptual situation. In the first instance it is an orientation with respect to one's own body, in this case one's perceptual apparatus. But once again we find that this first-order orientation shapes and structures our experience of our environment: *ta oikeia*, in this instance, are the appropriate objects of our perceptual apparatus and their proper 'fit' with our own perceptual equipment.

With Hierocles' conception of synaesthesia in hand, we can now return to the proto-transcendental claims we encountered in Cicero and Seneca. As we have seen, Seneca claims that self-consciousness is necessary, since it is somehow implicated in an

⁴⁷ In preparing the present chapter for press I learned of the recent publication of the first complete English edition of the extant texts of Hierocles, including the Greek text, an extensive bibliography, and supporting critical apparatus. Ramelli and Konstan (2009); for the passages quoted here, see p. 5, and critical commentary at pp. 39ff.

animal's awareness of other objects. Drawing on Hierocles' remarks about perception, we can now begin to recognize the motivation for this claim. Perception of an object, on this account, is to be understood as a form of endeavor or accomplishment—a form of 'agency', as Gummere renders the point in translating Seneca's letter. If this endeavor is not normally a salient aspect of a perceptual encounter, this can be put down to our extraordinary proficiency at carrying out everyday perceptual tasks. But as Hierocles observes, perceptual endeavor ('straining') does become salient in those cases where perceptual success is not immediate. The crucial point here is that in undertaking perceptual endeavors the organism does not simply register so much perceptual input; we will misunderstand the place of the perceiver if we think of our perceptual organs as simply on the receiving end of a causal chain. Rather, we must *exercise* our perceptual organs, bringing them to bear and attuning them to the perceptual object. But in order for this to be possible, we must have some tacit understanding of what those organs are, what they are for, and how they can be deployed so as to attain perceptual success.

If putting the point in this way runs the danger of making perception sound like a piece of intentional and strategic behavior, then we need only recall Seneca's insistence that this self-comprehension is an incipient and inarticulate native endowment of animal nature. Observe the blackbird in the garden, startled by a potential predator. Notice the way it cocks its head, first one way and then another, bringing the potential threat into the scope of binocular vision, assessing its size and distance and reacting accordingly. It is only in virtue of its ability to carry out this sequence of movements that the bird achieves perceptual success—gauging the threat and evading it. Its ability to do all this depends on its finely tuned prior orientation in its perceptual space, which in turn is the fruit of a intelligent mastery—instinctive and implicit to be sure—of the proper deployment of its perceptual apparatus. It is in this sense, I submit, that Seneca and Hierocles hold that self-consciousness (of the distinctively Stoic variety) serves as a condition on the possibility of objective perception.

What about Cato's claim in *De Finibus*? Does this analysis of synæsthetic perception generalize to the case of desire? The analogy is strained, I think, if only because we don't normally think of animals as having 'organs of desire' in quite the same way that they have perceptual organs. But like perception, desire is an intentional state: that is, both desire and perception require some thing or state-of-affairs as an object. The very possibility of desire, in this sense, presupposes some way in which its object can be or become determinate. Simply to be in a stew of dissatisfied affect does not amount to a state of desire any more than a show of colors and sounds itself amounts to perception of an object. So what is the role of self-consciousness in underwriting the possibility of desire? The key lies in understanding the distinctive structure and function of Stoic self-consciousness, and particularly the role it plays in providing an organism with normative orientation in its environment. To desire or 'have an appetite' for something presupposes that we have some way of distinguishing between what seems worth pursuing and what is to be avoided—predator from prey, nourishment from toxin, even pleasure from pain. Without some such capacity for discrimination desire could have

no determinate content and could provide no guidance in action. On the Stoic view, such distinctions are essentially kind-relative, hence the capacity to draw them must be grounded in an incipient understanding of one's distinctive mode of being. As we have seen, it is just such a self-understanding that the Stoics see as the central deliverance of self-consciousness.

6. Objections

There are no doubt many grounds on which one might object to the foregoing. Some objections pertain to the adequacy (or inadequacy) of the Stoic proofs that we have considered. An Epicurean might respond to Cato's argument, for instance, by granting the need for normative orientation in desire, but denying that this requires any kind of self-conscious self-comprehension or Hierocletian synaesthesia. Couldn't the sheer encounter with pleasure and pain—and thence with the objects that produce pleasure and pain—suffice to provide the requisite normative orientation? Call this *the Epicurean Retort*. A second line of objection might challenge the Stoic thesis that this form of primitive self-awareness amounts to comprehension of the constitution of the organism. An organism may be sufficiently oriented in its environment to be capable of desire even though it badly misunderstands its own bodily constitution. The ugly duckling is a swan but thinks it is a duck, and so does duckish things, whether or not duckish things benefit or befit it. So self-comprehending self-consciousness is not required for normative orientation. Call this *the Ugly Duckling Objection*.

Yet a third line of argument focuses on the logic of the Stoic arguments. As applied to human beings, the argument I have reconstructed effectively works from three premises. Very roughly:

- (i) Humans experience determinate desire and perceive determinate objects.
- (ii) In order to experience determinate desire and perceive determinate objects an organism must have normative orientation in its sphere of endeavor.
- (iii) Stoic self-consciousness provides normative orientation in an organism's sphere of endeavor.

From this it is concluded that humans have Stoic self-consciousness—an incipient, nondiscursive understanding of their own constitution and what befits it. But this looks to be a form of affirming the consequent. It neglects the possibility that there could be some other way in which normative orientation might be delivered. The Epicurean Retort proposes one alternate account of the source of that orientation, but even if the Epicurean alternative is found wanting, the Stoic proof cannot suffice to establish the transcendental thesis unless it can be shown that the requisite orientation could *only* be provided by Stoic self-consciousness.

A proper discussion of these and related objections must await another occasion, but I wish to conclude by considering one objection that is directed specifically to my thesis

that these Stoic arguments are aptly characterized as *transcendental*. If we take Kant as the gold standard for transcendental philosophy, then a number of fundamental differences immediately emerge. The Stoic arguments are not a priori, involving as they do rich and detailed empirical observations about human and animal behavior. They deal with psychological conditions, which Kantians (particularly neo-Kantians) have traditionally contrasted with genuinely transcendental constraints. And they do not in any straightforward way make what Mark Sacks once dubbed *the inference to reality*. The Stoic arguments advance claims *about our psychological condition*, rather than mounting any claim 'about how the world beyond the merely psychological realm must be'.⁴⁸

But here it is crucial to distinguish between the general form of a transcendental investigation and the particular dialectical purposes to which transcendental argumentation has been put. Kant himself sought to use transcendental argument in the first instance to answer a particular opponent: the Humean skeptic about causality. Latter-day Kantians have tried to use transcendental proof to defuse skepticism about the external world. If these are the primary deployments we have in mind for transcendental philosophy, then it is indeed crucial that a transcendental argument be wholly a priori and independent of any contingent psychological claims. One cannot use causal premises to answer a causal skeptic, nor empirical data to show that empirical knowledge is possible.

But it is important to recognize that the same general pattern of proof can be used to answer a different kind of opponent. The Stoics showed no particular concern with the kinds of skepticism that exercised Kant and his contemporaries, and their ultimate aims with the doctrine of *oikeiosis* were ethical rather than epistemic or metaphysical. The dialectical opponent against whom their proto-transcendental arguments were directed was the proponent of Epicurean moral psychology. Given those particular purposes and opponents, there is no threat of circularity in a transcendental argument that deploys empirical or causal premises. Moreover, they had no particular reason to insist that the conclusions of their arguments must concern 'a world beyond the merely psychological realm'. After all, the doctrine of *oikeiosis* was in the first instance a psychological theory. What the Stoics needed was a way of arguing from an undeniable premise to a contentious conclusion; the strategy they hit upon was to argue that the latter serves as a condition on the very possibility of the former. If my analysis is correct, their modal argument turned on a claim about the conditions under which a psychological state could obtain the proper 'fit' with its object. Whether we decide to call their arguments transcendental is perhaps in the end a strictly terminological matter; what seems clear is that the Stoics here developed a strategy of proof that was to have a considerable (and contentious) subsequent history.⁴⁹

⁴⁸ Sacks 2000, 273; see also Sacks 1999.

⁴⁹ Earlier versions of this chapter were presented at the University of Essex, New York University, the University of Sheffield, the University of York, the University of Newcastle, the History of Philosophy Roundtable at the University of California San Diego, and the University of Exeter, and at conferences of the International Society for Phenomenological Studies and the British Society for the History of Philosophy. I am grateful to the many friends and colleagues—by now too numerous to count—who have helped me with this material.

References

- Annas, Julia (1992), *Hellenistic Philosophy of Mind* (Berkeley, CA: University of California Press).
- Balibar, Étienne (1998), *John Locke, identité et différence: L'Invention de la conscience* (Paris: Seuil).
- Berges, Sandrine (2005), 'Loneliness and Becoming: Is Stoic Cosmopolitanism Still Defensible?', *Res Publica*, 11/1, 3–25.
- Brink, C. O. (1956), 'Oikeiosis and Oikeiotes: Theophrastus and Zeno on Nature in Moral Theory', *Phronesis*, 1, 123–45.
- Brunschwig, J. (1986), 'The Cradle Argument in Epicureanism and Stoicism', in M. Schofield and G. Striker (eds), *The Norms of Nature: Studies in Hellenistic Ethics* (Cambridge: Cambridge University Press), 113–45.
- Cole, Jonathan (1995), *Pride and a Daily Marathon* (Cambridge, MA: MIT Press/Bradford Books).
- Grotius, Hugo (1625), *Prolegomena to the Law of War and Peace*; citations refer to translation by F. W. Kelsey (Indianapolis: Bobbs-Merrill, 1957).
- Hill, L. (2000), 'The Two Republicae of the Roman Stoics: Can a Cosmopolite be a Patriot?', *Citizenship Studies*, 4/1, 65–79.
- Irwin, Terence (2003), 'Stoic Naturalism and its Critics', in B. Inwood (ed.), *The Cambridge Companion to Stoicism* (Cambridge: Cambridge University Press), 345–64.
- Long, A. A., and D. N. Sedley (1987), *The Hellenistic Philosophers* (Cambridge: Cambridge University Press).
- Long, A. A. (1993), 'Hierocles on Oikeiosis and Self-Perception', in K. Boudouris (ed.), *Hellenistic Philosophy* (Athens: International Association For Greek Philosophy), 93–104; citations refer to reprint in A. A. Long, *Stoic Studies* (Berkeley, CA: University of California Press, 1996), 250–63.
- Martin, Wayne (2010), 'The Judgment of Adam: Self-Consciousness and Normative Orientation in Lucas Cranach's Eden', in Joseph Parry (ed.), *Art and Phenomenology* (London: Routledge), 105–37.
- Nussbaum, Martha (1997), 'Kant and Stoic Cosmopolitanism', *Journal of Political Philosophy*, 5/1, 1–25.
- Nussbaum, Martha (2000), 'Duties of Justice, Duties of Material Aid: Cicero's Problematic Legacy', *Journal of Political Philosophy*, 8/2, 176–206.
- Pagden, A. (2000), 'Stoicism, Cosmopolitanism, and the Legacy of European Imperialism', *Constellations*, 7/1, 3–22.
- Pangle, T. (1998), 'Socratic Cosmopolitanism: Cicero's Critique and Transformation of the Stoic Ideal', *Canadian Journal of Political Science/Revue Canadienne de Science Politique*, 31/2, 235–62.
- Pembroke, S. G. (1971), 'Oikeiosis', in A. A. Long (ed.), *Problems in Stoicism* (London: Athlone), 114–59.
- Pohlenz, Max (1940), 'Grundfragen der stoischen Philosophie', *Abhandlungen der Gesellschaft der Wissenschaften zu Göttingen*, 3rd ser. 26, 1–103.

- Ramelli, Ilaria, and David Konstan (2009), *Hierocles the Stoic: Elements of Ethics, Fragments, and Excerpts* (Atlanta, GA: Society of Biblical Literature).
- Sacks, Mark (1999), 'Transcendental Arguments and the Inference to Reality: A Reply to Stern', in R. Stern (ed.), *Transcendental Arguments: Problems and Prospects* (Oxford: Oxford University Press, 1999), 67–82.
- Sacks, Mark (2000), *Objectivity and Insight* (Oxford: Oxford University Press).
- Sherrington, Charles (1906), *The Integrative Action of the Nervous System* (Cambridge: Cambridge University Press).
- Schofield, Malcolm (2003), 'Stoic Ethics', in B. Inwood (ed.), *The Cambridge Companion to the Stoics* (Cambridge: Cambridge University Press, 2003), 233–56.
- Striker, Gisela (1983), 'The Role of *Oikeiosis* in Stoic Ethics', *Oxford Studies in Ancient Philosophy*, 1, 145–67.
- Striker, Gisela (1991), 'Following Nature', *Oxford Studies in Ancient Philosophy*, 9, 1–73.
- von Armin, H. (1926), 'Arius Didymus' Abriss der peripatetischen Ethik', *Sitzungsberichte der Akademie Wien*, 204/3.
- Whewell, William, translator (1853), *Grotius on the Rights of War and Peace: An Abridged Translation* (London: John Parker).

Index

- a priority 16, 40, 65, 71, 77, 197–8, 282, 286, 287, 313n39
- cognition 9, 21–3, 24, 26, 74, 75, 86, 136–7
- concepts 21–3, 218–19
- consciousness 176, 177
- Fichte 78, 80, 86, 139
- genetic descriptions 78, 80
- Heidegger 279–80, 284–5, 286
- Husserl 234
- intuitions 75–6
- judgements 21n3, 205, 218
- Kant 21–3, 24, 26, 35, 36–7, 38, 71, 74, 75–6, 77, 91, 136–7, 147, 150, 174–5, 234, 246
- knowledge 23, 35, 36–7, 38, 137, 197n4, 279–80, 284, 286
- incommensurability of 288–92
- Merleau-Ponty 300, 319n39
- Nietzsche 195, 199, 200, 210–11, 212, 213, 216–19, 220, 223, 224–5
- propositions 91, 136
- respect 147, 150
- Sartre 296–7
- synthesis of apprehension 174–5
- absolute activity 143: *see also* freedom
- absolute freedom 189
- absolute idealism 159–72
- absolute knowing 185
 - Fichte 125, 126–9, 131
- abstraction 80–1, 84–5
 - abstractive reflection 108–9, 114
- actions 164–5
- activity 72, 81, 140–2
 - absolute activity 143
 - and consciousness 79, 93, 108, 151
 - Fichte 109, 110, 111, 112, 119, 120, 122, 123, 124, 127, 129, 130, 138, 139, 140–1, 149–50
 - of I 109, 110, 111, 112, 119, 120, 122, 123, 124, 129, 139
 - ideal activity 140–1
 - real activity 140–1
 - self-activity 144–5, 146, 148–9, 155
 - self-reverting activity 80, 140–1
- Aenesidemus* (Schulze) 102
- Allison, Henry E. 40, 46
- ambiguity 315–18
- animal psychology 358
- animals, self-consciousness of 355–61
- Annas, Julia 344, 362
- Anscombe, Elizabeth 142
- anti-naturalism 303
- antinomies 312–15, 317
 - antinomy of pure reason 6n7, 30–4, 312, 313, 314
- appearances 10, 11, 48, 222
 - cognition and 6
 - Hegel 179
 - Heidegger 273
 - Husserl 258
 - Kant 24–5, 26, 33, 40, 43–5, 47, 49–50, 51–2, 100–1
 - Leibniz 28n15
 - Merleau-Ponty 313–14
 - Müller-Lyer lines 167, 315–16
 - Nietzsche 196, 204, 221
 - and representations 24, 33, 51–2
 - and things in themselves 49, 50, 100–1, 204
 - Williams 331–2
- apperception
 - apperceptive judgements 161
 - transcendental apperception 175–6, 177
 - unity of 176, 177, 180–1, 187, 248
- apprehension 230, 248, 311
 - synthesis of 165, 174–5
- Aristotle 200, 269, 272n13, 345, 348
- assertions 107–8, 265–8, 271, 272–3, 274–7
- assumption 284
- autonomy 58, 146, 248, 280
- Baldwin, Thomas 299
- Basic Problems of Phenomenology*, *The* (Heidegger) 287
- Basic Questions of Philosophy* (Heidegger) 290
- Beck, Jakob Sigismund 13–14
- Beck, Lewis White 68n18
- behaviour
 - human 57n1, 58, 64, 65–6, 72
 - moral 59
- behavioural economics 62–3
- behaviourism 356n34
- being
 - being-in-the-world 249–50, 265–6, 271, 276, 277
 - being-uncovering 265–6, 268, 270–1, 272, 275, 276, 277
 - and freedom 128, 141
 - Sartre 296–7
 - see also* *Being and Time* (Heidegger); *Dasein*
- Being and Nothingness* (Sartre) 296–7
- Being and Time* (Heidegger) 229, 267, 275, 278–92

- assertion/belief 276n
- Dasein 278, 281–5
- detranscendentalizing interpretations
 - of 280–3
- entities 283–4, 285–6, 287–8, 289, 290
- hermeneutic idealism 283–5
- hermeneutic reading of 279–80
- hermeneutic transformation of
 - transcendental philosophy 280–3
- incommensurability thesis 288–92
- truth 265–7, 273
 - understanding, fore-structure of 285–8
- being-in-the-world 249–50, 265–6, 271, 276, 277
- being-uncovering 265–6, 268, 270–1, 272, 275, 276, 277
- beliefs 42, 168, 307, 311
 - justified true beliefs 5, 8, 103–4
 - Nietzsche 217, 218–19, 225
 - truth and 268, 272–7
- Berkeley, George 15n
- Berlin Phenomenology, The* (Hegel) 163
- Beyond Good and Evil* (Nietzsche) 214–15, 217, 218, 221n41, 224, 264
- Bible
 - Job, Book of 71–2
 - Pauline epistles 346–7
- Briefe über die Kantische Philosophie* (Reinhold) 137n
- Brunschwig, J. 347–52
- Bubner, Rüdiger 247
- Candide* (Voltaire) 69
- Carman, Taylor 228–9, 282n14
- Cartesianische Meditationen* (Husserl) 231, 240
- Cassirer, Ernst 9–10, 15n
- categorical imperative 56, 58, 64, 66
 - transcendental deduction of 136, 140–7
- categories 113n, 211, 257–8, 259, 287
 - Fichte 132
 - Hegel 165, 187, 193
 - Kant 23, 25–6, 30, 32, 175, 247, 255–6, 288–9
 - Merleau-Ponty 316
 - neo-Kantian 248–9, 255
 - Nietzsche 212, 219, 222
 - representations and 26n11
- Cato the Younger 352–5, 361–2, 364
- causality, principle of 26n12
- causes: and reasons 64, 66
- Cavell, Stanley 324, 325, 335, 339–40
- Cherniss, H. 344
- child psychology 358
- Chrysippus 345, 346, 359
- Cicero, Marcus Tullius 344, 352–3, 354, 359, 361–2, 364
- Clark, Maudemarie 195, 196, 197n4, 198n6, 199n8, 215, 217
- cognition 2n1, 12
 - a priori 9, 21–3, 24, 26, 74, 75, 86, 136–7
 - and appearances 6
 - empirical 75–6
 - genetic 89
 - Hegel 178–9
 - mathematical 75
 - philosophical 75
 - synthetic 75–6
- Cohen, Hermann 248–9
- Collier, Arthur 30n
- colors 11n13, 48, 52n44, 258, 333–4
- commitment 262–3, 276–7
- common sense 304, 305
- common sense realism 196, 225n
- communication 206
- concept determinacy 166n
- concepts 165–6
 - a priori concepts 21–3, 218–19
 - conscious/unconscious 209
 - and intuition 160, 161, 162, 165
 - Kant 175, 218–19
- conceptual pluralism 285, 288, 292
- conceptual self-determination 167–8
- conceptual unboundedness 160, 167–8
- conditioned propositions 112, 116–17
- confirmation holism 290
- consciousness 172, 185–7
 - a priori 176, 177
 - and activity 79, 93, 108, 151
 - Fichte 79, 80, 81, 82, 87, 88, 91n65, 93, 142, 149
 - and freedom 141
 - Hegel 172, 185–6, 187
 - Husserl 251–2
 - of identity 175
 - and intentionality 190, 259
 - Kant 175
 - and meaning 258–9
 - ordinary consciousness 178–80, 181–3, 185, 187–8, 190
 - and self-consciousness 148, 151, 187–8
 - and sense-certainty 190–2
 - spirit and 188–9, 192
 - of time 174–5
 - truth and 182
 - of unity 175–6
- constructivism 332, 334, 336
- Copernican revolution
 - Heidegger on 281, 284
 - Kant 1, 2, 4–6, 11, 14, 35, 230, 284, 300, 305, 306
 - Merleau-Ponty 300, 306
- correctness 277
 - definition of 269–70
 - Heidegger 268–9, 270–1
 - truth as 268–70
- correspondence: truth as 265–9, 270, 271–2
- cosmopolitanism 347–52
- Cox, Christoph 195, 222

- Critique of Practical Reason* (Kant) 60–1, 62, 67, 68, 136–7, 147, 150
- Critique of Pure Reason* (Kant) 74, 205
- categories 25–6
 - imagination 202
 - ontology 23, 26
 - synthetic method 75
 - 'transcendental', definition of 21–4
 - transcendental deduction in A edition 174–7
 - transcendental deduction in B edition 176, 177
 - transcendental idealism 24–5
 - transcendental realism 24, 25
- Critique of the Power of Judgement* (Kant) 70, 72, 90n63
- Crowell, Steven 279, 280–2
- Dasein 261–3, 265–6, 268, 270–1, 277, 278, 281–5: *see also* being
- Davidson, Donald 276n
- De Finibus* (Cicero) 352–3, 354, 361–2, 364
- De Man, Paul 199n8
- deductions 91–2
- demonstrations 87, 88, 90–2, 93–4
- Dennett, Daniel 57, 59, 234
- dependence 166–7, 169
- Descartes, René
- a priori knowledge 197n4
 - self-consciousness, Cartesian view of 355–6, 358, 359
- desire 258
- and self-consciousness 361–2, 364–5
- determination 166–7
- Fichte 141–2
 - Hegel 165
 - self-determination 141–2
- differentiated intuition 174–5
- discontinuity thesis 246, 253
- divisibility 113
- Diogenes Laertius 345, 354
- Diogenes the Cynic 347
- Dreyfus, Herbert 251n26
- duty 58, 60n6, 63n11, 64, 150–1, 206–7
- Dwyer, Philip 228
- Eilan, Naomi 296n
- Elements of Ethics* (Hierocles) 362–3
- Eleutheriologie, or on Freedom and Necessity* (Ulrich) 137n
- Eluard, Paul 201
- empirical cognition 75–6
 - empirical explanations 302–3, 329, 330–1, 336
 - empirical idealism 332
 - empirical psychology 318, 319
 - empirical realism 27, 34
 - empiricism 317
 - Merleau-Ponty 301, 302–3, 305–6
- Engstrom, Stephen 166–7
- entities 283–4, 285–6, 287–8, 289, 290
- Epicurean psychology 352, 353, 354
- Epicureanism: and Stoic
- transcendentalism 352–5
- Erste Philosophie I* (Husserl) 230
- Erste Philosophie II* (Husserl) 232, 236
- Essence of Truth, The* (Heidegger) 266–7, 269, 273, 274–5
- ethical universalism 348
- evolutionary psychology 66, 217
- experience 208–12, 251–5
- Nietzsche 195, 197–8, 199, 200, 202–4, 208–13
- explanations 8, 11, 12, 57–8, 253–4, 290, 300–1
- empirical 302–3, 329, 330–1, 336
 - genetic 89, 90, 105, 109, 116, 128, 143
 - of human behaviour 65–6, 156, 159
 - Kant 12, 72, 101
 - Merleau-Ponty 297, 306, 307
 - transcendental 306, 308
- 'Faith and Knowledge' (Hegel) 180–1
- falsification thesis 205, 207–8, 220, 221–2
- feelings 136–7, 147–52
- moral feelings 147–52
 - respect 147, 150
 - sympathetic feeling 150–1
- Fichte, J. G.
- absolute knowing 125, 126–9, 131
 - abstraction 80–1, 84–5
 - abstractive reflection 108–9, 114
 - activity 109, 110, 111, 112, 119, 120, 122, 123, 124, 127, 129, 130, 138, 139, 140–1, 149–50
 - anti-skeptical strategies 106–32
 - categories 132
 - conditioned/unconditioned propositions 112
 - consciousness 79, 80, 81, 82, 87, 88, 91n65, 93, 142, 149
 - divisibility 113
 - freedom 88n56, 128, 139, 141–2, 143, 145–6, 147, 148–9, 152–8
 - I as Act 109, 111–12, 114, 117–20, 121–3
 - I-hood 128, 139, 155–6
 - I/non-I 108–14, 117–18, 149
 - intellectual intuition 121–3, 148
 - intuitions 91, 92, 121–3, 148
 - judgements 124
 - knowing 129–31
 - knowledge, foundationalist theory of 107
 - knowledge, grounding-oriented theories of 106, 107–13, 132
 - knowledge, justification-oriented theory of 105, 106
 - knowledge, phenomenological approach to 125–6
 - knowledge, self-consciousness and theory of 99, 108, 114–15

- Fichte, J. G. (*Cont.*)
 knowledge, skeptical challenge to argument
 for 115–17, 123–4
 moral failure 154
 moral feeling, transcendental argument
 for 147–52
 moral law, deduction of 140–7
 natural/moral drives 154–6
 propositions 107–11, 112, 116–17, 124
 psychology 152
Reellephilosophie 81–2, 84–5
 reflection 80, 81–2, 83, 84, 87–8
 representations 81, 82, 83, 90, 138–9, 141,
 142–3
 self-awareness 156–7
 self-consciousness 79–80, 88, 99, 108, 114–15,
 139, 151, 153, 156–7
 self-observation 84, 90
 synthetic-genetic method 78–94
 transcendental ethics 135–58
 unconditioned, concept of 107–8, 110–12, 113,
 115–17, 119, 146
 willing and self-activity 144–5, 146
 final justifications 2
 finite idealism 159–72
 first-person plural idealism 329
 fore-conception 287
 fore-having 287
 fore-sight 287
 Foucault, Michel 224, 240–1
Foundations of Natural Right (Fichte) 79, 81, 85
Foundations of the Entire Wissenschaftslehre
 (Fichte) 85, 87, 107
Foundations of Transcendental Philosophy
 (*Wissenschaftslehre*) *nova methodo* (Fichte)
 81, 82, 87, 117, 139–42
 activity 140–1
 representations 138
 self-consciousness 139
 and self-determination 141–2
 freedom
 absolute freedom 189
 and being 128, 141
 and consciousness 141
 and Dasein 262
 Fichte 88n56, 128, 139, 141–2, 143, 145–6, 147,
 148–9, 152–8
 and generality 146
 Hegel 188
 Kant 50n39, 137, 153
 Merleau-Ponty 308, 310
 and moral law 137
 noumenal freedom 47n28
 problem of 152–8
 as representation of self-activity 143
 and self-consciousness 139, 155, 157–8
see also absolute activity
- Friedländer, Paul 270
 Friedman, Michael 161
 functional state identity theory 356n34
- Gadamer, Hans-Georg 164
Gay Science, The (Nietzsche) 215–16, 219, 221
 generality 2, 146
 genetic cognition 89
 genetic descriptions 78, 80
 genetic explanations 89, 90, 105, 109, 116, 128,
 143
 Genova, A. C. 245n
 geometry 38, 89–93
 German idealism 9, 96–103
 Gettier problem 205n19
 givenness 32, 119, 166, 233, 252–3, 258
 God 11–12, 50n39, 68, 71, 170
 good life 56–7, 60–1, 62, 67
 Greek rationalism 168
 Grotius, Hugo 344
Groundwork of the Metaphysics of Morals (Kant)
 59, 60, 62, 136, 137, 153
 Gummere, Richard M. 356n36, 357, 364
 Guyer, Paul 8n, 64n13, 68n18
- Hacker, Peter 324, 325
 Hales, Steven 222n
 happiness 43, 56, 70
 and naturalism 59–64
 and practical reason 59–64, 67
 and virtue 60, 70
 Hartmann, Klaus 159, 249–50, 254
 Haugeland, John 262
 Hegel, G. W. F. 103, 159–72, 173–94
 appearances 179
 categories 165, 187, 193
 cognition 178–9
 conceptual self-determination 167–8
 consciousness 172, 185–6, 187
 determination 165
 freedom 188
 identification 165
 imagination 180–1
 inseparability 171–2
 intuitions/concepts 160, 161, 165–6
 judgements 127n42
 knowledge, grounding-oriented theories
 of 106
 life/love concepts 120n
 metaphysics 45, 159–60
 negation 163
 normativity 166
 ontology 193
 phenomenology 173–4, 185–9
 quasi-transcendental thought 178–84
 reason/Greek rationalism 168
 sense-certainty 183–4, 187, 188, 189–94

- spatiotemporal predicates 45
- speculative philosophy 178–81
- and transcendental philosophy 173–4, 178–84
- Heidegger, Martin 11n12, 228–9, 247n10, 286n22, 356n34
- and a priority 279–80, 284–5, 286
- appearances 273
- assertion 276n
- assumptions 284
- being-in-the-world 276
- being-uncovering 265–6, 268, 270–1, 272–3
- belief 276n
- commitment 262–3
- consciousness/care 259
- Copernican revolution 281, 284
- correctness 268–9, 270–1
- correspondence 265–9
- Dasein 261–3, 265–6, 268, 270–1, 277, 278, 281–5
- hermeneutic idealism 283–5
- hermeneutics 279–85
- incommensurability thesis 288–92
- and Kant 249
- representations 267–8
- subjectivity 260
- transcendence 250–1, 260–2
- transcendental philosophy, hermeneutic transformation of 280–3
- truth 265–8, 272, 273
- unconcealment 264–77
- see also *Being and Time*
- Henrich, Dieter 97n, 119n31, 245
- Herman, Barbara 59
- hermeneutic idealism 283–5, 291
- hermeneutics
 - conceptual pluralism 285
 - Heidegger and 279–85
 - incommensurability thesis 285
 - transcendental philosophy, transformation of 280–3
- Hicks, Robert Drew 345, 346
- Hierocles 347nn19& 20, 349–50, 362–3, 364
- highest good
 - Kant 150
 - and naturalism/ practical reason 60–3, 67–8
- History of the Concept of Time* (Heidegger) 286n22
- Hölderlin, Friedrich 127&n42
- holism
 - circle of understanding 286–7, 291
 - confirmation holism 290
 - ontological holism 44–5
- Homer 270
- Horwich, Paul 272
- Human, All Too Human* (Nietzsche) 214, 215, 221
- human behaviour 57n1, 58, 64, 72
 - explanation of 65–6, 156, 159
- human psychology 150, 348
- Hume, David 5, 202
- Husserl, Edmund 220, 221, 228–41
 - and a priority 234
 - adequation 271
 - appearances 258
 - consciousness 251–2, 259n50
 - intentionality 250–5
 - intersubjectivity, problem of 234–9
 - and Kant 230, 234–9
 - meaning 252
 - noema 252
 - normality 237–9
 - objectivity 236, 237, 238
 - phenomenology, assessment of 228–9, 230
 - phenomenology, transcendental nature of 230–1
 - psychology 234
 - self-evidence 271
 - subjectivity 259n50
 - transcendence 231, 235, 250
 - transcendental idealism 231
 - transcendental phenomenology 229, 230–3
 - transcendental reduction 229, 230–3
 - truth 271, 272
 - validity claim 260
- Hyppolite, Jean 299–300
- I-hood 108–14, 117–23, 128, 139, 155–6
 - I as Act 109, 111–12, 114, 117–20, 121–3
 - I/non-I 108–14, 117–18, 149
 - Kantian theory 120–1
 - self-positing activity 109, 110, 111, 112, 119, 120, 122, 123, 124, 129, 139
 - see also self-consciousness
- ideal activity 140–1
- idealism 22n6, 253
 - empirical idealism 332
 - finite/absolute idealism 159–72
 - first-person plural idealism 329
 - German idealism 9, 96–103
 - hermeneutic idealism 283–5, 291
 - Merleau-Ponty 298, 312–15
 - psychological idealism 40, 50
 - Wittgenstein 327, 329, 332
 - see also transcendental idealism
- Ideen II* (Husserl) 237
- identification 165
- identity 123, 169, 181, 286, 313, 350n
 - consciousness of 175–6, 192
 - functional state identity theory 356n34
 - identity interpretation 40, 49–50
 - identity theory 160
 - logical law of 109–10, 112
- illusion 44, 167, 223, 335
 - transcendental illusion 31–3, 335

- imagination 15n, 86–7, 209
 Hegel 180–1
 Kant 25n, 165, 175, 180–1, 202
 Nietzsche 201–3, 206, 208, 209, 211
 productive imagination 180–1
 Sartre 296
 Wittgenstein 330, 339, 340
 see also metaphors
- immediacy 101, 104, 119, 122, 168, 191, 256
 immorality 152–7
 incommensurability thesis 285, 288–92
 indeterminacy 315–16, 317
 individuation 38–9, 284, 287–8
 inseparability claim 171–2
 intellectual intuitions 121–3, 148
 intellectualism 161, 301, 317
 intentionality 190, 250–5, 257, 259, 260
 intersubjectivity 234–9, 310
 intuitions 75–6, 92, 165–6
 a priori 75–6
 and absolute knowing 128
 and concepts 160, 161, 162, 165
 differentiated intuition 174–5
 Fichte 91, 92, 121–3, 148
 intellectual intuitions 121–3, 148
 Kant 165, 176
- Jacobi, F. H. 39, 45, 100–2
 Job, Book of 71–2
 judgements 63, 107, 127, 339–40
 a priori judgments 21n3, 205, 218
 apperceptive judgements 161
 Cassirer 10, 15n
 Fichte 124
 Kant 9, 10, 120
 Merleau-Ponty 304, 316
 moral judgements 58, 65–6
 Nietzsche 200, 202, 217, 218
 practical judgements 70
 justified true beliefs 5, 8, 103–4
- Kant, Immanuel 77, 17
 and a priority 21–3, 24, 26, 35, 36–7, 38, 71, 74, 75–6, 77, 91, 136–7, 147, 150, 174–5, 234, 246
 anti-naturalist reading 57–9
 antinomy of pure reason 312, 313, 314
 appearances 24–5, 26, 33, 40, 43–5, 47, 49–50, 51–2, 100–1
 apperception, synthetic unity of 176, 180–1, 187
 apprehension, synthesis of 165, 174–5
 categorical imperative 136, 144
 categories 23, 25–6, 30, 32, 175, 247, 255–6, 288–9
 causality, principle of 26n12
 concepts 165, 175
 consciousness 175
 Copernican revolution 1, 2, 4–6, 11, 14, 35, 230, 284, 300, 305, 306
 demonstration, concept of 90
 discontinuity thesis 246
 explanations 12, 72, 101
 and feelings 136–7, 147, 150–1
 freedom 50n39, 137, 153
 happiness 67
 and Heidegger 249
 highest good 150
 Husserl and 230, 234–9
 I, theory of 120–1
 identity 175
 imagination 25n, 165, 175, 180–1, 202
 indeterminacy 315n31
 intentionality 252
 intuitions 165, 176
 judgements 9, 10, 120
 knowledge, theory of 108
 and Merleau-Ponty 313
 moral law 136–8
 phenomenal/noumenal selves 137–8
 possibility, conditions of 2
 predispositions 152
 propositions 26n12, 75–6
 psychology 150, 248–9
 representation 22–3, 24, 26n11, 29n16, 33, 51–2, 101–2, 121, 174–5, 176, 177, 187, 246–7
 reproduction, synthesis of 175
 respect 136, 147, 150
 self-consciousness 108
 skepticism 100–2
 things in themselves 24–5, 26, 27, 30, 35–6, 37, 38–9, 41–3, 46, 47–9, 50, 52–3, 100–1, 204, 312
 transcendental aesthetic 5
 transcendental analytic 5–6
 transcendental deduction 136, 174–7, 245–6, 247–8
 transcendental idealism 21–6, 51–2, 312
 transcendental realism 24–6, 46–7
 unconditioned, concept of 43, 46, 47–8, 70, 101–2
 and Wittgenstein 324, 325, 335, 340
 see also Critique of Practical Reason; Critique of Pure Reason; transcendental idealism Kant und die Idee der Transzendentalphilosophie (Husserl) 234
- King Lear* (Shakespeare) 72
 knowing 51n41, 115–16, 121, 217n31, 335
 absolute knowing 125, 126–9, 185
 Fichte 99–100, 110, 115, 117, 119, 120, 122, 125–32
 Hegel 185
 knowing subject 2, 113, 114, 125, 130, 247
 relative (particular) knowing 126, 131
 theory of 98, 99–100

- knowledge 197, 204–5
 a priori 23, 35, 36–7, 38, 137, 197n4, 279–80, 284, 286
 capacity for 170
 Fichte 125–6
 foundations of 2
 Gettier problem 205n19
 and justified true beliefs 103–4
 Nietzsche 196, 198&n5, 199–200, 204, 205, 207, 214–15, 217, 218, 219, 221, 223–6
 objective knowledge 196, 198&n5, 199–200
 phenomenological approach 125–6
 and skepticism 98–106, 115–17, 123–4
- knowledge, theories of 98–113
 contextualist 105, 106
 foundationalist 105–6, 107
 German idealism and 98–103
 grounding-oriented 105–6
 justification-oriented 105–6, 107–13, 132
 relativistic 105–6
 and self-consciousness 99, 108, 114–15
- Kofman, Sarah 199n8
- Krause, Karl Christian Friedrich 140
Krisis der europäischen Wissenschaften, Die (Husserl) 232, 236, 239
- Langton, Rae 47–8
- language 169, 201, 218–19, 267–8, 329–31
 language games 335, 338
 limits of 168, 170–1, 325, 329, 333, 336
 of negation 164, 165
- Lask, Emil 255–7, 260
- Lear, Jonathan 170
- Lectures on Logic* (Kant) 77
- Lectures on the Doctrine of Religion* (Kant) 61
- Leibniz, Gottfried Wilhelm 28n15, 197n4
- Leiter, Brian 195, 196n3, 199, 217, 225n
- Lisbon earthquake (1755) 69
- logical empiricism 281n11
- Logische Untersuchungen* (Husserl) 234
- Long, A. A. 344, 346n17, 353, 356, 357, 360
- McDowell, John 160, 164n7, 167–8, 247n8, 254
- Maimon, Salomon 101n5
- Marburg School 248–9
- Martin, Gottfried 23n8
- mathematical cognition 75
- mathematics 38, 89–93
- meaning 251–4, 261, 263, 286–7
 and consciousness 258–9
 constructivist theory of 332, 334, 336
 Lask 256–7
 and reference 286, 290
- Mendelssohn, Moses 11n13
- Merleau-Ponty, Maurice 229
 a priori 300, 319n39
 ambiguity 315–18
- antinomies 312–15, 317
- appearances 313–14
- categories 316
- Copernican revolution 300, 306
- empiricism 301, 302–3, 305–6
- explanations 297, 306, 307
- freedom 308, 310
- idealism 298, 312–15
- intellectualism 301, 317
- intersubjectivity 310
- judgements 304, 316
- and Kant 313
- on Müller-Lyer lines 315–16
- naturalism 299, 302, 306, 309, 318
- objectivity 304
- perception, theory of 294–321
- perception, theory of: psychological interpretation 295–300, 301n14, 309, 310, 312, 320
- perception, theory of: transcendental interpretation 300–2, 307, 310, 320
- perception, theory of: and transcendental metaphysics 302–12
- pre-objectivity 298, 299, 301–2, 304–7, 308–10, 311n26, 312, 313–15, 316–17
- presuppositions 304–5
- psychology 18, 317n, 318–20
- representations 298, 315
- and Schelling 320n40
- solipsism 314–15
- temporality 300
- values 309–10
- metaphors 200–2
- Aristotle 200
- Nietzsche 201–2, 206, 207, 208–10, 211
- Metaphysical Foundations of Logic, The* (Heidegger) 283n
- metaphysical realism 283, 287
- metaphysics
- Hegel 45, 159–60
- and transcendental idealism 36–7
- and transcendental turn 2
- ‘Metaphysics and the Novel’ (Merleau-Ponty) 311n27
- Metaphysics of Morals* (Kant) 63–4
- freedom, problem of 153
- happiness 67
- sympathetic feeling 150–1
- Moore, A. W. 171
- Moore, G. E. 277
- moral agency 57
- moral behaviour 59
- moral drives 154–6
- moral failure 154
- moral feeling: transcendental argument for 147–52
- moral judgements 58, 65–6

- moral law 136–8, 146
 deduction of 140–7
 phenomenal/noumenal 137–8
 moral psychology 152, 352
 Moran, Dermot 239
Mots et les choses, Les (Foucault) 240
 Müller-Lyer lines 167, 315–16
- Nachgelassene Fragmente* (Nietzsche) 212, 216,
 218, 223–4
 natural drives 154–6
 natural realism 231
 naturalism 16, 56–73, 197, 319–20
 and happiness 59–64
 and highest good 60–3, 67–8
 Merleau-Ponty 299, 302, 306, 309, 318
 Nietzsche 196, 197n4
 and obligation 63–7
 scientific naturalism 309
 and tragedy 67–72
 transcendental anti-naturalism 303
 working description of 57–9
 negation 5, 11n13, 113n24, 163
 language of 164, 165
 self-negation 162, 163–5, 166–7
 Nehamas, Alexander 217
 Nietzsche, Friedrich 195–226
 a priority 195, 199, 200, 210–11, 212, 213,
 216–19, 220, 223, 224–5
 appearances 196, 204, 221
 and belief 217, 218–19, 225
 categories 212, 219, 222
 duty 206–7
 error theory 196, 198, 204–5, 207, 216, 217–20,
 222–5
 experience 195, 197–8, 199, 200, 202–4,
 208–13
 falsification thesis 205, 207–8, 220, 221–2
 imagination 201–3, 206, 208, 209, 211
 judgements 200, 202, 217, 218
 knowledge 196, 198&n5, 199–200, 204, 205,
 207, 214–15, 217, 218, 219, 221, 223–6
 metaphors 201–2, 206, 207, 208–10
 naturalism 196, 197n4
 perception 197–8, 200, 201–2, 208–11, 219
 perspectivism 216–17, 225n
 representations 200, 202, 203, 206, 211, 223
 and science 195–6
 sense data 220–1
 things in themselves 196–8, 200, 204, 205,
 211, 215, 216, 224
 transcendental idealism 195–200
 truth 264
 world of becoming 221, 222, 223
 noema 252–3, 257–9
 normality 166, 237–9
 normativity 166, 168, 171–2, 260–3
- noumenal freedom 47n28
 Nussbaum, Martha 350n
- objective knowledge 196, 198&n5, 199–200
 objectivity
 Husserl 236, 237, 238
 Merleau-Ponty 304
 objective and subjective 142, 143, 148
 see also pre-objectivity
 obligation 56, 57, 63–7
 naturalism and 63–7
 practical reason and 63–7
 reasons and 64–5
oikeiosis 342–66
 animal self-consciousness 355–61
 concentric circles image 347, 349
 Epicureanism 352–5
 interpretation of 343–7
 and Stoic cosmopolitanism 347–52
 'On Instinct in Animals' (Seneca) 356–60, 362
On the Genealogy of Morals (Nietzsche) 225n
On the Miscarriage of All Philosophical Trials
 (Kant) 70, 71
 ontic transcendence 250–1
 ontological holism 44–5
 ontology
 Hegel 193
 Kant on 23, 26
 and spatiotemporal predicates 24–5, 27–30
 Wolff on 22n4
 ordinary consciousness 178–80, 181–3, 185,
 187–8, 190
 original (primal) transcendence 250–1, 260–3
 particular (relative) knowing 126, 131
- Pauline epistles 346–7
 Pembroke, S. G. 344, 345
 perception 257–8
 Merleau-Pontian theory of 294–321
 Merleau-Pontian theory of: psychological
 interpretation 295–300, 301n14, 309, 310,
 312, 320
 Merleau-Pontian theory of: transcendental
 interpretation 300–2, 307, 310, 320
 Merleau-Pontian theory of: and
 transcendental metaphysics 302–12
 Nietzsche 197–8, 200, 201–2, 208–11, 219
 and self-consciousness 362–4
 and sense-certainty 190, 191
 perspectivism 216–17, 225n
Phänomenologie und Anthropologie (Husserl)
 232
Phenomenological Interpretations of Kant's
'Critique of Pure Reason' (Heidegger) 284
 phenomenology
 Hegel 173–4, 185–9
 Husserl 228–41

- in Kantian context 245–50
- and phenomenologies 244–5
- and transcendental philosophy 244–63
- Phenomenology of Perception* (Merleau-Ponty)
 - ambiguity 315–18
 - antinomy 312–15, 317
 - idealism 312–15
 - perception, theory of: psychological
 - interpretation of 295–300, 301n14, 309, 310, 312, 320
 - perception, theory of: transcendental
 - interpretation of 300–2, 307, 310, 320
 - perception, theory of: transcendental metaphysics and 302–12
 - psychology 317–18
- Phenomenology of Spirit* (Hegel) 162, 166, 171–2, 173–94
 - consciousness 185
 - (quasi-)transcendental thought 178–84
- 'Philosopher, The' 202, 203n17, 204, 205, 210, 214
- philosophical cognition 75
- Philosophy of Right* (Hegel) 165
- physicalism 57, 65
- Pinder, Tilmann 21n3
- Pippin, Robert 173–4, 179n14, 189–90, 192–3
- Platner, Ernst 83
- Plato 269, 270, 273, 348
- pluralism 285
- Plutarch 347–8
- Poellner, Peter 215, 217n31
- Pohlenz, Max 343n3
- possibility, conditions of 2, 7–12
- practical intentionality 260
- practical judgements 70
- practical reason 56–73
 - and happiness 59–64, 67
 - and highest good 60–3, 67–8
 - naturalism, working description of 57–9
 - and obligation 63–7
 - and tragedy 67–72
- pre-objectivity 298, 299, 301–2, 304–7, 308–10, 311n26, 312, 313–15, 316–17
- predispositions 152
- presuppositions 304–5
- Priest, Stephen 300n12
- 'Primacy of Perception and its Philosophical Consequences, The' (Merleau-Ponty) 300, 307, 310, 313, 314
- primal (original) transcendence 250–1, 260–3
- productive imagination 180–1
- productivity 119
- Prolegomena* (Kant) 22, 51–2, 74
- propositions 80n27, 86, 94, 157, 246
 - a priori 91, 136
 - conditioned/unconditioned propositions 112, 116–17
- Fichte 107–11, 112, 116–17, 124
- Kant 26n12, 75–6
 - synthetic propositions 26n12, 75–6, 86, 136
 - transcendental propositions 76
 - and truth 267–8
 - see also* assertions; judgements
- proprioception 360–1
- Prose of the World, The* (Merleau-Ponty) 311n25
- psychoanalytic theory 356n34
- psychological egoism 345–6, 350
- psychological idealism 40, 50
- psychologism 80n28, 255–60, 319
- psychology 2, 26n12, 65, 68n18, 77, 97, 202, 318–21, 347, 358
 - empirical psychology 318, 319
 - Epicurean psychology 352, 353, 354
 - evolutionary psychology 66, 217
 - human psychology 150, 348
 - Kant 150, 248–9
 - Merleau-Ponty 18, 317n, 318–20
 - moral psychology 152, 352
 - and perception, theory of 295–300, 301n14, 309, 310, 312, 320
 - psychological egoism 345–6, 350
 - and representation 254
 - scientific psychology 18, 303, 319
 - Stoic psychology 351–2, 358, 361–2, 366
 - transcendental psychology 35–6, 37–8
- psychological idealism 40, 50
- quasi-transcendental thought 178–84
- Rackham, Harris 344n5, 353, 362
- rationalism 168
- real activity 140–1
- realism 251
 - common sense realism 196, 225n
 - metaphysical realism 283, 287
 - natural realism 231
- reason 168: *see also* practical reason
- reasons 64–5, 66
- reductive physicalism 65
- reference: and meaning 286, 290
- reflection
 - abstractive reflection 108–9, 114
 - Fichte 80, 81–2, 83, 84, 87–8
- reflexivity 119
- Reinhold, K. L. 102, 137n
- relative (particular) knowing 126, 131
- Religion within the Boundaries of Mere Reason* (Kant) 153, 137n
- representations 15n, 42n17, 49–50, 90, 167, 169
 - and appearances 24, 33, 51–2
 - and categories 26n11
 - Fichte 81, 82, 83, 90, 138–9, 141, 142–3
 - Heidegger 267–8
 - Kant 22–3, 24, 26n11, 29n16, 33, 51–2, 101–2, 121, 174–5, 176, 177, 187, 246–7

- representations (*Cont.*)
 Merleau-Ponty 298, 315
 Nietzsche 200, 202, 203, 206, 211, 223
 and psychology 254
 representationalism, critique of 252–5
 spatiotemporal 291n16, 211
 Wittgenstein 334
 reproduction, synthesis of 175
Republic (Plato) 269
 respect 136, 147, 150
 Rickert, Heinrich 254
 Rorty, Richard 170, 198n6
 Rousseau, Jean-Jacques 69
 Ryle, Gilbert 356n34

 Sacks, Mark 211–12, 257n46, 290, 366
 Sartre, Jean-Paul 296–7, 310
 Scheler, Max 233
 Schelling, Friedrich 99, 106, 320n40
 schematism 202
 Schlösser, Ulrich 126n
 Schmid, C. E. 82, 137n
 Schofield, Malcolm 344
 Schulze, G. E. 42n17, 102
 science 195–6
Science of Logic (Hegel) 178, 179, 193
 scientific naturalism 309
 scientific psychology 18, 303, 319
 Scotus, John Duns 111n13
 Searle, John 251n26
 Sedley, D. N. 346n17, 357
 self-activity 144–5, 146, 148–9, 155
 self-awareness 156–7: *see also*
 self-consciousness
 self-consciousness 2n1, 97–8, 142–3, 151–2, 260
 of animals 355–61
 Cartesian view 355–6, 358, 359
 and consciousness 148, 151, 187–8
 and desire 361–2, 364–5
 Fichte 79–80, 88, 139, 142, 151, 153, 156–7
 Fichte and theory of knowledge 99, 108,
 114–15
 and freedom 139, 155, 157–8
 and perception 362–4
 and proprioception 360–1
 Stoics/Seneca 355–61
 and willing 146
 see also I-hood; self-awareness; subjectivity
 self-determination 141–2, 167–8
 self-negation 162, 163–5, 166–7
 self-observations 84, 90
 self-referentiality 247, 248
 self-reverting activity 80, 140–1
 self-understanding 260
 Sellars, Wilfrid 161
 semantic externalism 356n34
 Seneca 355–61, 362, 363–4

 sense-certainty
 and consciousness 190–2
 Hegel 183–4, 187, 188, 189–94
 and perception 190, 191
 sense data 220–1
 sensibility 160
 Shakespeare, William 72
 Sherrington, Charles 360
 skepticism 2, 5–6
 Hume on 5
 and knowledge 98–106
 knowledge, challenge to Fichtean argument
 for 115–17, 123–4
 and truth 178
 Socrates 270, 273
 solipsism
 Merleau-Ponty 314–15
 truth in 168, 170
 Wittgenstein 168, 170, 326–8
 souls 68
 spatiotemporal predicates 31–2, 36–7, 38, 48, 211
 German romantics and 44–5
 Kant on 24–5, 27–30
 speculative philosophy 178–81
 Spinoza, Baruch 45, 72, 163
 spirit: and consciousness 188–9, 192
 Stack, George 199n7, 217
 Stoic cosmopolitanism 347–52
 Stoic psychology 351–2, 358, 361–2, 366
 Stoic transcendentalism 342–66
 Epicureanism 352–5
 interpretation of *oikeiosis* 343–7
 objections to 365–6
 and Stoic cosmopolitanism 347–52
 Stoics: and self-consciousness 355–61
 Strasser, Stephan 236
 Strawson, P. F. 170–1, 197
 Striker, Gisela 351n28
Structure of Behaviour, The (Merleau-Ponty)
 296, 303n
 subjectivism 49–50, 142, 143, 148
 subjectivity 2n1, 97, 255–60: *see also*
 self-consciousness
 sympathetic feeling 150–1
 synaesthesia 363
 synthetic cognition 75–6
 synthetic propositions 26n12, 75–6, 86, 136
System of Ethics (Fichte) 81, 88, 89, 135, 140, 142–4
 activity 149–50
 consciousness 142, 149
 freedom 145–6, 154, 155, 156
 self-consciousness 139, 142

 Taylor, Charles 173–4, 183–4
 temporality 300
Theaetetus (Plato) 269, 270
 theodicy 69, 71–2

- things in themselves
 - and appearances 49, 50, 100–1, 204
 - existence of 43–4, 196, 197
 - Kant 24–5, 26, 27, 30, 35–6, 37, 38–9, 41–3, 46, 47–9, 50, 52–3, 100–1, 204, 312
 - Nietzsche 196–8, 200, 204, 205, 211, 215, 216, 224
 - unknown 52–3
 - Wittgenstein 335
- thought
 - quasi-transcendental 178–84
 - self-negation of 162
- time: consciousness of 174–5
- Tractatus Logico-Philosophicus* (Wittgenstein) 168–9, 326–8
- tragedy 67–73
- transcendence
 - Heidegger 250–1, 260–2
 - Husserl 231, 235, 250
 - ontic transcendence 250–1
 - original (primal) transcendence 250–1, 260–3
- transcendental aesthetic 5, 27–30
- transcendental analysis 66
- transcendental analytic 5–6
- transcendental anti-naturalism 303
- transcendental apperception 175–6, 177: *see also* self-consciousness
- transcendental argumentation 2
- transcendental constraints 290
- transcendental deduction 719, 25n
 - of categorical imperative 136, 140–7
 - in *Critique of Pure Reason* A edition 174–7
 - in *Critique of Pure Reason* B edition 176, 177
 - Kant 136, 174–7, 245–6, 247–8
- transcendental dialectic 46–7
- transcendental ethics 135–58
 - categorical imperative, transcendental deduction of 136, 140–7
 - challenges for 135–40
 - freedom, problem of 152–8
 - moral feeling, transcendental argument for 147–52
- transcendental explanations 306, 308
- transcendental features 290
- transcendental idealism 4–7, 14–15, 196–7, 281, 313–14
 - causality interpretation 40, 49–50
 - context of 39–40
 - and empirical realism 34
 - Garve-Feder/Göttingen Review 22
 - Husserl 231
 - identity interpretation 40, 49–50
 - Kant 21–6, 51–2, 312
 - and metaphysics 36–7
 - moderate interpretation 36–9, 40–9, 53
 - Nietzsche 195–200
 - and spatiotemporal predicates 29
 - standpoint 13
 - and transcendental illusion 32–3
 - Williams 325–7, 330, 332, 333, 335–7, 339
- transcendental illusion 31–3, 335
- transcendental logic 2
- transcendental metaphysics 302–12
- transcendental phenomenology 229, 230–3
- transcendental philosophy
 - Heidegger's hermeneutic transformation of 280–3
 - mixed theories 249
 - and phenomenology 244–63
 - synthetic method 74–7
- transcendental proof 2
- transcendental propositions 76
- transcendental psychology 35–6, 37–8
- transcendental realism 14–15, 24–6, 28, 31–3, 46–7
- transcendental reality 46
- transcendental reduction 229, 230–3
- transcendental reflection 177, 181
- transcendental standpoint 13–16
- transcendental thought 178–84
- transcendental turn: definition 1–3
- truth 264–7
 - and belief 268, 272–7
 - and consciousness 182
 - as correctness 268–70
 - as correspondence 265–9, 270, 271–2
 - Heidegger 265–8, 272, 273
 - Plato 269, 273
 - and propositions 267–8
 - redundancy theory 272
 - skepticism and 178
 - in solipsism 168, 170
- 'Truth and Lies' (Nietzsche) 196, 198–200, 201–4, 205–7, 208, 209–11, 212, 220
- Tugendhat, Ernst 272–3, 274
- Ulrich, Johann August Heinrich 137n
- unboundedness 160, 161–2, 167–8
- unconcealment 264–77
- unconditioned, concept of 45, 47–8, 223
 - Fichte 107–8, 110–12, 113, 115–17, 119, 146
 - Kant 43, 46, 47–8, 70, 101–2
- unconditioned propositions 112, 116–17
- uncovering 275, 277
- understanding 160, 248, 282, 285–8
- unity: consciousness of 175–6
- universalism 348, 349
- values 309–10
- Versuch einer Moralphilosophie* (Schmid) 137n
- virtue 60, 70
- Visible and the Invisible, The* (Merleau-Ponty) 304–5, 310–11

- Voltaire (François-Marie Arouet) 69, 70
 von Arnim, H. 343n3
Vom Wesen der Wahrheit (Heidegger) 286n21

 Welshon, Rex 222n
What is a Thing? (Heidegger) 290
 Whewell, William 344
 Wieland, Wolfgang 186n
Will to Power (Nietzsche) 212, 215, 216n, 218n36,
 219, 220–1, 223
 Williams, Bernard 67n15, 169, 170
 appearances 331–2
 on early Wittgenstein 325–8
 on later Wittgenstein 328–32, 337
 on transcendental idealism 325–7, 330, 332,
 333, 335–7, 339
 on Wittgenstein 324–40
 willing 144–5, 146
Wissenschaftslehre
 anti-skeptical strategies 106–32
 objections to 127
 synthetic-genetic method 78–94
 'Wittgenstein and Idealism' (Williams) 324–40
 early Wittgenstein 325–8
 later Wittgenstein 328–32, 337
 Wittgenstein, Ludwig 168–70, 171
 constructivism 332, 334, 336
 first-person plural use 328–32, 333, 335–6,
 337–40
 idealism 327, 329, 332
 imagination 330, 339, 340
 and Kant 324, 325, 335, 340
 language-games 335, 338
 language, limits of 170
 and natural science 333–4
 representations 334
 solipsism 168, 170, 326–8
 things in themselves 335
 truth 168, 170, 275–6
 Williams on 324–40
 Williams on: early Wittgenstein 325–8
 Williams on: later Wittgenstein
 328–32, 337
 Wolff, Christian 22n4
 Wood, Allen 39–40, 49–53, 59
 Wrathall, Mark 266–8, 272, 274

 Zeno 347–8
 Zettel (Wittgenstein) 333, 334
 Zöller, Günter 125n39
Zur Phänomenologie der Intersubjektivität I
 (Husserl) 236
Zur Phänomenologie der Intersubjektivität III
 (Husserl) 236, 238
Zur phänomenologischen Reduktion: Texte aus
dem Nachlass (Husserl) 232–3
Zweite Einleitung in die Wissenschaftslehre
 (Fichte) 117

